

..... NNT/NL : 2020AIXM0366/028ED251

## THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université  
le 30 novembre 2020 par

**Benjamin Bourel**

**Utilisation des pollens pour la reconstruction à haute résolution spatiale des environnements à Hominini du Plio-Pléistocène dans le rift est-africain (Éthiopie et Tanzanie)**

**Pollen use for the spatial high-resolution reconstruction of Plio-Pleistocene Hominini environments in the East African Rift (Ethiopia and Tanzania)**

**Discipline**

Sciences de l'Environnement

**Spécialité**

Ecologie et Anthropologie Biologique

**École doctorale**

251

**Laboratoire de recherche**

Centre européen de Recherche et d'Enseignement de Géosciences de l'Environnement (CEREGE)



- **Composition du jury**
- **Odile PEYRON : Rapporteur**
- DR CNRS, ISEM, Université de Montpellier (France)
- **Sarah IVORY : Rapporteur**
- Assistant professor, Dpt. of Geosciences, Univ. of Penn State (USA)
- **Andrew COHEN : Examinateur**
- Professor, Dpt. of Geosciences, University of Arizona (USA)
- **Joel GUIOT: Examinateur**
- DR CNRS émérite, CEREGE, Université d'Aix-Marseille (France)
- **Raymonde BONNEFILLE : Membre invité**
- Ancien DR CNRS, CEREGE, Université d'Aix-Marseille (France)
- **Luc BEAUFORT : Directeur de thèse**
- DR CNRS, CEREGE, Université d'Aix-Marseille (France)
- **Doris BARBONI : Co-directeur de thèse**
- CR CNRS, CEREGE, Université d'Aix-Marseille (France)



*"Les enseignements qui se font sans douleur n'ont pas de réelle valeur. Ainsi, on ne peut rien obtenir sans faire de sacrifice."*

*Hiromu Arakawa*



## Affidavit

Je soussigné, Benjamin Bourel, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Doris Barboni et Luc Beaufort, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisées dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Aix-en-Provence, le 25 septembre 2020

I, undersigned, Benjamin Bourel, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Doris Barboni and Luc Beaufort, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Aix-en-Provence, date 25 September 2020



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](#).

## Résumé long

Ce mémoire de thèse par articles présente l'étude du signal pollinique de sédiments modernes, submodernes (760-2011 *Anno Domini*, AD) et de 3,42-2,96 millions d'années (Ma) (Pliocène) en Ethiopie et dans le nord de la Tanzanie dans le but de contribuer à une meilleure connaissance des micro-habitats associés aux Hominini du Plio-Pléistocène du rift est-africain. Le jeu de données modernes pour cette étude inclut 283 spectres polliniques de surface (sols et vases) dont 62 inédits provenant des formations arbustives à *Acacia-Commiphora* et de zones afromontagnardes qui étaient peu documentées. Il a été développé ici, par des analyses statistiques et de l'apprentissage automatique sur ce jeu de données modernes, une approche qui rend maintenant possible d'identifier 24 types et sous-types de végétations dont 9 de végétations locales ripariennes ou édaphiques. Cette calibration met aussi en avant l'importance de traiter différemment les données polliniques en fonction des conditions de dépôt : classes d'abondances pour les milieux de dépôt peu dynamiques (eg., sols ou bordures lacustres) et présence-absence pour les milieux de dépôt dynamiques (eg., lits ou bordures de rivières). Appliquée aux spectres polliniques fossiles des formations pliocènes de Hadar et Busidima (basse vallée de l'Awash, Éthiopie), cette nouvelle approche a permis d'interpréter/réinterpréter de façon plus objective des données anciennes (33 spectres) et nouvelles (9 spectres provenant du sondage HSPDP-NAW-1A). Les résultats montrent qu'entre 3,42 Ma et 2,96 Ma, le Nord Afar a subi d'importants refroidissements associés aux événements globaux des stades isotropiques marins M2, MG2 et MG4 (-6,5°C à -10,5°C par rapport à l'actuel) qui ont induit une succession d'environnements très variés à l'échelle régionale durant cette période, alors qu'au même moment, localement, la paléorivière Awash et les paléosystèmes de sources d'eau souterraine ont permis le maintien de formations boisées ou forestières au niveau des sites à *Australopithecus afarensis*, leur offrant ainsi un habitat boisé stable. Au vu de l'importance écologique des résurgences en zone aride et le nombre de sites paléoanthropologiques (>50) associés à des travertins déposés en contexte de sources, il a été décidé d'étudier un sondage provenant d'un marécage à *Typha* (Kisima Ngoda, bordure nord du Lac Eyasi, Tanzanie) alimenté par un système de sources jugé analogue à ceux de certains sites paléoanthropologiques. L'enregistrement pollinique

montre que les habitats boisés à *Acacia* et palmiers *Hyphaene* sont présents au moins depuis 1250 ans, que la résurgence répond aux changements de précipitations régionales, mais qu'un système de double aquifère permet aux micro-habitats boisés de se maintenir à Kisima Ngeda même pendant des périodes arides. Les reconstructions de paléovégétations associées aux Hominini étant largement limitées par le grand nombre d'échantillons sub-stériles en pollens, il a aussi été décidé de travailler sur un système de reconnaissance automatique. Basé sur des réseaux de neurones convolutifs, le système mis au point ici est capable d'identifier avec succès (taux d'erreur <4%) des grains de pollens modernes et fossiles, et ce, quel que soit leur état de dégradation, ce qui est une avancée majeure. En conclusion, cette thèse contribue à ce que les pollens fossiles permettent de caractériser les paléoenviro-nnements avec un degré de précision loin d'être à ce jour égalé par d'autres proxys. Notre étude révèle l'importance des micro-habitats boisés associés aux Hominini du Pliocène, dont l'écologie n'est pas encore clairement établie. Elle pourrait contribuer à rejeter l'hypothèse de la savane. Malgré les faibles taux de préservation des pollens dans les niveaux pliocènes, la reconnaissance automatique dont nous avons démontré ici le potentiel pourra probablement aider dans un futur proche à reconstituer avec une meilleure résolution spatiale l'habitat des ancêtres de la lignée humaine.

**Mots clés :** paléo-environnement, Pliocène, Hadar, Olduvai, *Australopithecus*, réseau de neurones convolutifs

## Long abstract

This thesis by articles aims at contributing to a better understanding of the micro-habitats associated with the Plio-Pleistocene Hominini of the East African Rift, by means of a study of the pollen signal of modern, submodern (760-2011 *Anno Domini*, AD) and 3.42-2.96 million years (Ma) (Pliocene) sediments in Ethiopia and northern Tanzania. The modern dataset collected in this study uses 283 surface pollen samples (soils and mud) including 62 previously unpublished samples from *Acacia-Commiphora* bushlands and Afromontane zones that were previously poorly documented. In this modern dataset, 24 types and sub-types of vegetation, including 9 of local riparian or edaphic vegetation were identified by applying statistical analyses and machine learning. This calibration highlights the importance of treating pollen data in dependence of the deposition conditions: abundance classes for less or no dynamic depositional environments (eg., soils or lake edges) and presence-absence for dynamic depositional environments (eg., river beds or edges). Applied to the fossil pollen spectra of the Pliocene Hadar and Busidima Formations (Lower Awash Valley, Ethiopia), this new approach allowed a more objective interpretation/reinterpretation of old (33 spectra) and new (9 spectra from HSPDP-NAW-1A) data. Results show that between 3.42 Ma and 2.96 Ma, North Afar experienced significant cooling associated with global events of the marine isotopic stages M2, MG2 and MG4 (-6.5°C to -10.5°C compared to the present). The climatic and vegetation reconstructions induce a succession of very varied environments on a regional scale during this period, while at the same time, locally, the Awash paleoriver and the paleosystems of groundwater sources supported wooded or forest formations at the sites of *Australopithecus afarensis*, thus providing them with a stable wooded habitat. In view of the ecological importance of dryland springs and the number of paleoanthropological sites (>50) associated with tufa deposited in the context of springs, it was subsequently decided to study a core from a *Typha* swamp (Kisima Ngeda, northern edge of Lake Eyasi Tanzania) fed by a spring system, considered analogous to those of some paleoanthropological sites. The pollen record shows that (i) *Acacia* and *Hyphaene* Palm woodland habitats have been present for at least 1250 years, (ii) the spring responds to regional rainfall changes, (iii) that a dual aquifer system allows the wooded micro-habitats to persist in Kisima Ngeda even during arid periods. Since the reconstructions of

paleovegetation associated with the Hominini are broadly limited by the large number of near sterile pollen samples, it was also decided to work on an automated recognition system. Based on convolutional neural networks, the system developed here is capable of successfully identifying (error rate <4%) modern and fossil pollen grains regardless of their state of degradation, which we consider to be a major advance in palynological studies. In conclusion, this thesis contributes to the fact that the analysis of fossil pollen allows the characterization of paleoenvironments with a degree of accuracy difficult to obtain with other proxies to date. Our study reveals the importance of the wooded micro-habitats associated with the Pliocene Hominini, whose ecology is to date not well established. It could contribute to rejecting the savannah hypothesis. In spite of the low rates of Pliocene pollen preservation, the automated recognition whose potential we have demonstrated here may help in the near future to reconstitute with a better spatial resolution the habitat of the ancestors of the human lineage.

**Keywords:** paleoenvironment, Pliocene, Hadar, Olduvai, *Australopithecus*, convolutional neural network

## **Remerciements**

En premier lieu, je tiens à remercier Guillaume BUCHET qui nous a malheureusement quitté au cours de cette thèse. Il a été d'une aide inestimable dans l'identification des pollens et le partage sans limite de son savoir avec moi. C'était un homme passionné et d'une grande gentillesse. J'aurais aimé lui montrer tout ce que j'ai pu accomplir grâce à lui.

Je remercie mes directeurs de thèse, Doris BARBONI et Luc BEAUFORT qui chacun à leur façon, m'ont permis de mener à bien ce travail. Merci à eux d'avoir cru en moi et de m'avoir donné l'opportunité de réaliser cette thèse. Leur investissement et leur soutien malgré les difficultés rencontrées ont été remarquables. À Doris, merci d'avoir partagé avec moi tes connaissances et de m'avoir motivé et inspiré lorsque j'en avais besoin. Je te remercie pour ton aide et ta bienveillance. Tu as toujours fait de ton mieux pour me mettre en avant et m'aider à réaliser mes projets. Tu m'as soutenu et aidé à poursuivre mes recherches même lorsque ces dernières n'allait pas dans le sens de tes paradigmes. Tu es une chercheuse comme il en faudrait plus.

Je remercie également les autres membres du jury. Andrew COHEN et Joël GUIOT pour leur participation au jury de thèse et Odile PEYRON et Sarah IVORY pour être les rapporteurs de ce travail. Merci à Raymonde BONNEFILLE d'avoir accepté de se joindre à nous en tant que membre invité et pour ces conseils tout au long de la thèse.

J'adresse mes remerciements aux membres du CEREGE pour leur accueil, leur aide et leurs encouragements. Je pense notamment à Anne ALEXANDRE, Françoise CHALIÉ, Thibault DE GARIDEL-THORON, Yves GALLY, Gislain GASSIER, Joël GUIOT, Guillaume LEDUC, Christine PAILLÈS et Baptiste SUCHÉRAS-MARX. Merci à Didier ZEVACO pour ses nombreux coups de main en informatique, mais surtout pour son amitié et sa sympathie. Un grand merci à Jean-Charles MAZUR et Philippe DUSSOUILLEZ pour leur aide et leur investissement dans ma thèse ainsi que pour leur bonne humeur et leur gentillesse. Également merci à Olivier BELLIER, Noémie GARRON et Sylvie DE FREITAS pour leur aide et leur soutien dans l'ensemble des démarches administratives réalisées durant cette

thèse. Enfin, merci à Isabelle HAMMAD pour sa disponibilité, son assistance et sa gentillesse.

Je remercie aussi tous les extérieurs qui ont rendu ce travail possible avec une pensée pour Anne-Marie LÉZINE pour m'avoir aidé à accéder aux données de l'*African Pollen Database*. Et aussi Morteza DJAMALI et Suzanne LEROY pour leurs aides et conseils sur les analyses polliniques, mais aussi leur bonne humeur et leur sympathie.

Bien évidemment, je remercie tous mes amis doctorants et post-doctorants pour leur bonne humeur, les pauses café, les goûters de l'ACET , les soirées et les sorties que nous avons faites ensemble ! Une pensée particulière pour Irène AUBERT, Nith CAM, Riccardo CATALANO, Amandine JEAN, Nicolas LAUNAY, Brice LEBRUN (best co-bureau), Nicolas LUSINIER, Maximilien MATHIAN, Romain MUNCK, Alice NOVELLO, Marie PROTIN, Solène QUERO, Clément OUTREQUIN. Enfin, un très grand merci pour mes amis Jonathan COUTAZ et Jean-David MOREAU qui m'ont toujours aidé et soutenu même dans les moments les plus difficiles.

Pour terminer, un très grand merci également à toute ma famille pour leur soutien sans faille, leurs nombreux encouragements et leur gentillesse. Merci à eux d'avoir cru en moi. Je n'aurais jamais pu accomplir tout ceci sans eux.

# Table des matières

AFFIDAVIT .....	1
RESUME LONG .....	2
LONG ABSTRACT .....	4
REMERCIEMENTS.....	6
TABLE DES MATIERES .....	8
1. INTRODUCTION GENERALE.....	11
1.1. CADRE ET OBJECTIFS DE LA THESE.....	11
1.2. PRESENTATION DES ETUDES .....	13
1.3. DEVELOPPEMENTS ET VALORISATIONS ANNEXES .....	20
2. INFERENCE DES TYPES DE VEGETATION LOCALE ET REGIONALE A PARTIR DES ASSEMBLAGES POLLINIQUES : UNE NOUVELLE APPROCHE POUR ETUDIER LA PALEOECOLOGIE DES PREMIERS HOMININI .....	21
2.1. INTRODUCTION .....	22
2.2. STATE OF THE ART.....	25
2.3. STUDY AREA .....	27
2.4. MATERIAL .....	32
2.4.1. <i>Modern pollen samples and sampled climate space</i> .....	32
2.4.2. <i>Geopositioning and assignment of environmental data</i> .....	33
2.5. ANALYTICAL APPROACH .....	35
2.5.1. <i>Original pollen data and transformations</i> .....	35
2.5.2. <i>Data mining through clustering and vegetation types identifiable using pollen assemblages</i> .....	36
2.5.3. <i>Vegetation score for the km-groups (V score)</i> .....	39
2.5.4. <i>Vegetation score in dynamic depositional environment (V-river score)</i> .....	41
2.5.5. <i>Significant and non-significant score ranges for V and V-river scores</i> .....	41
2.5.1. <i>Environmental parameters for post-hoc inferred vegetation</i> .....	41
2.5.2. <i>Rainfall patterns for post-hoc inferred vegetation</i> .....	42
2.6. RESULTS.....	43
2.6.1. <i>Comparison of km-groups with the real distribution of local and regional vegetation at the sampling sites</i> .....	43
2.6.2. <i>Effectiveness of V and V-river scores in vegetation identification</i> .....	44
2.6.3. <i>Relation between inferred vegetation types and sub-types and environmental parameters</i> .....	46
2.6.4. <i>Rainfall paterns</i> .....	51
2.6.5. <i>Morphologies and vegetation strip of rivers</i> .....	53
2.6.6. <i>Most characteristic pollen types</i> .....	54
2.7. DISCUSSION .....	56
2.7.1. <i>Usefulness for environmental reconstructions</i> .....	56
2.7.2. <i>A more objective approach</i> .....	60
2.8. CONCLUSIONS .....	61
2.9. ACKNOWLEDGMENTS.....	62
3. AUSTRALOPITHECUS AFARENSIS RELIE AUX ENVIRONNEMENTS BOISES : RECONSTRUCTION PALYNOLOGIQUE A HAUTE RESOLUTION SPATIALE POUR LES PALEO-VEGETATIONS DES FORMATIONS DE HADAR ET BUSIDIMA (AFAR, ETHIOPIE).....	63

3.1. INTRODUCTION .....	64
3.2. MATERIAL AND METHOD .....	68
3.3. RESULTS.....	75
3.3.1. <i>Stratigraphic position and age of the pollen samples</i> .....	75
3.3.2. <i>Reconstruction of paleovegetation and paleoclimate</i> .....	79
3.4. DISCUSSION .....	83
3.4.1. <i>Importance of reinterpreting old pollen samples</i> .....	83
3.4.2. <i>Climatic reconstruction and evidence of MIS M2, MG2, and MG4</i> .....	86
3.4.3. <i>Variation of Intertropical Convergence Zone (ITCZ)</i> .....	88
3.4.4. <i>The ecology of Australopithecus afarensis in the Northern Awash Basin.</i> .....	89
3.5. CONCLUSIONS .....	91
3.6. ACKNOWLEDGMENTS.....	92
<b>4. ENREGISTREMENT POLLINIQUE ET DYNAMIQUE DES MICROHABITATS ASSOCIES A LA RESURGENCE D'EAU DOUCE SOUTERRAINE DE KISIMA NGEDA, LAC EYASI, TANZANIE .....</b>	<b>93</b>
4.1. INTRODUCTION .....	94
4.2. STUDY AREA .....	98
4.3. MATERIAL AND METHOD .....	102
4.4. RESULTS.....	105
4.4.1. <i><sup>14</sup>C dating and age models</i> .....	105
4.4.2. <i>Pollen preservation and representation</i> .....	105
4.4.3. <i>Three successive pollen zones indicate increasing groundwater influence since AD 1200.</i> .....	108
4.5. DISCUSSION .....	110
4.5.1. <i>Microhabitats changes and groundwater level</i> .....	110
4.5.2. <i>Comparison with other regional records</i> .....	113
4.5.3. <i>Spring vegetation response to IOD positive events</i> .....	114
4.5.4. <i>Acacia-Hyphaene palm woodland versus spring forest</i> .....	115
4.5.5. <i>Kisima Ngeda, a dual and therefore sustainable groundwater system?</i> .....	117
4.5.6. <i>Impact of springs on the geological and paleobotanical records at hominin sites</i> .....	118
4.6. CONCLUSIONS .....	119
4.7. ACKNOWLEDGMENTS.....	120
<b>5. RECONNAISSANCE AUTOMATISEE DE GRAINS DE POLLEN MODERNES, FOSSILES, INTACTS ET ENDOMMAGES PAR RESEAUX NEURONNAUX CONVOLUTIFS MULTIPLES.....</b>	<b>121</b>
5.1. INTRODUCTION .....	122
5.2. MATERIAL AND METHOD .....	126
5.2.1. <i>Datasets and image acquisition</i> .....	126
5.2.2. <i>Image classification</i> .....	128
5.3. RESULTS.....	131
5.3.1. <i>Global comparison of used/unused augmentation and single-CNNs/ multi-CNNs</i> .....	131
5.3.2. <i>Details of efficiency of Multi-CNNs with augmentation</i> .....	132
5.4. DISCUSSION .....	135
5.5. CONCLUSIONS .....	138
5.6. COMPUTER CODE AVAILABILITY .....	140
5.7. ACKNOWLEDGMENTS.....	140
<b>6. SYNTHESE.....</b>	<b>142</b>
<b>BIBLIOGRAPHIE .....</b>	<b>146</b>
<b>TABLE DES FIGURES .....</b>	<b>161</b>
<b>TABLE DES TABLEAUX .....</b>	<b>165</b>
<b>TABLE DES ANNEXES .....</b>	<b>166</b>

<b>ANNEXE DES FIGURES .....</b>	<b>168</b>
<b>ANNEXE DES TABLEAUX.....</b>	<b>181</b>
<b>ANNEXE DES PUBLICATIONS .....</b>	<b>194</b>

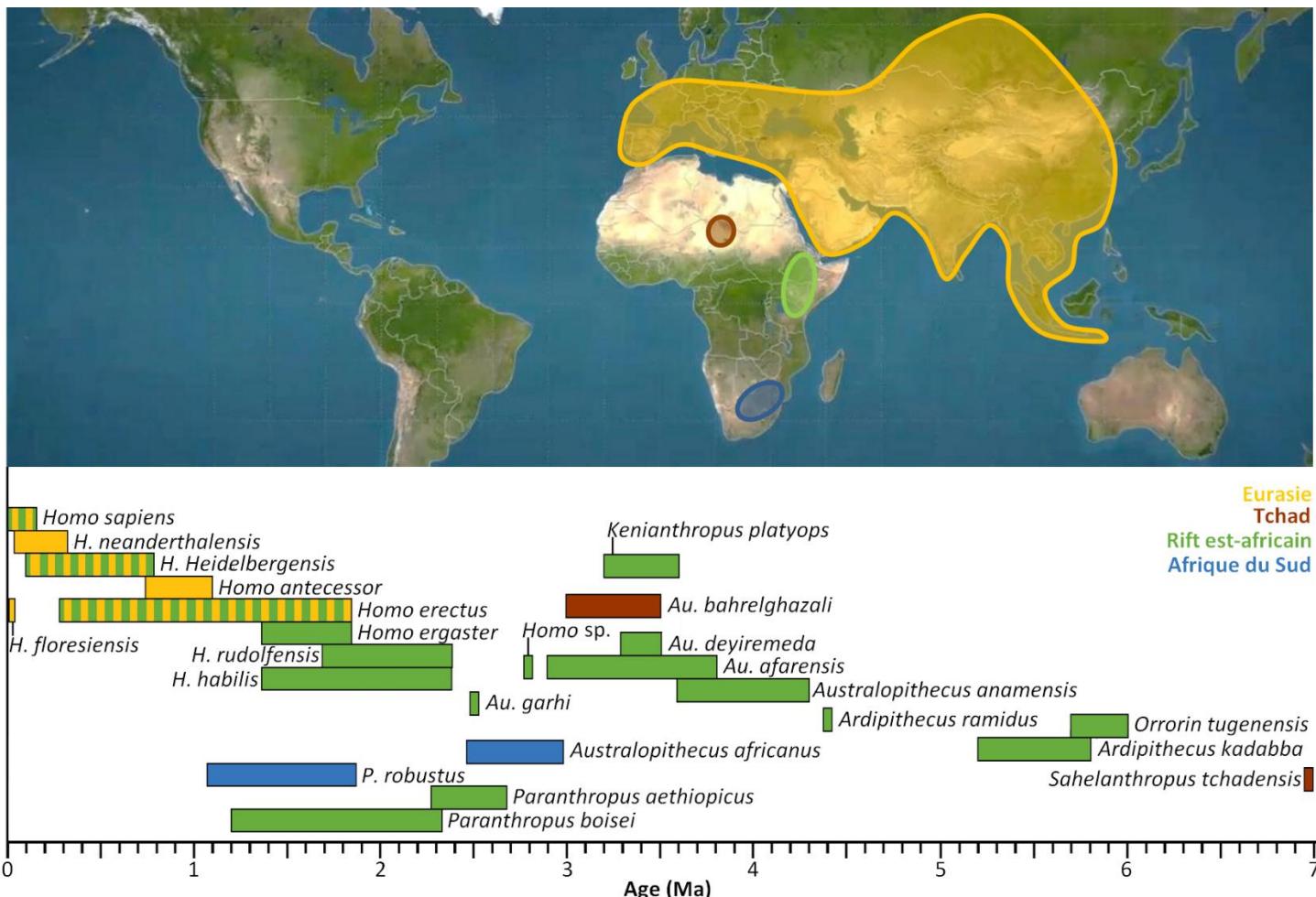
# 1. Introduction générale

## 1.1. Cadre et objectifs de la thèse

Le rift est africain est dans sa branche Est (Ethiopie, Kenya et Nord de la Tanzanie) une zone riche en fossiles d'Hominini<sup>a</sup> (**Figure 1.1**). Ils y sont attestés depuis au moins 6 millions d'années (Ma) avec les restes d'*Orrorin tugenensis* de 6,0 Ma à 5,7 Ma (collines de Tugen, Kenya) puis de ceux d'*Ardipithecus kadabba* de 5,8 Ma à 5,2 Ma (moyenne vallée de l'Awash, Ethiopie) et d'*Ardipithecus ramidus* à 4,4 Ma (moyenne vallée de l'Awash, Ethiopie) (Haile-Selassie et al., 2004; Richmond and Jungers, 2008; White et al., 2009). Il s'en est suivi dans cette zone géographique des événements de spéciation qui ont mis en place la première espèce d'*Australopithecus* connue avec *Australopithecus anamensis* de 4,3 Ma à 3,6 Ma (Ethiopie et Kenya) (Bobe et al., 2020; White et al., 2006). A partir de là, une diversification importante des Hominini va commencer à s'opérer dans cette partie du rift avec l'apparition de nouvelles espèces d'*Australopithecus* mais aussi la mise en place des genres *Paranthropus* (première occurrence vers environ 2,7 Ma) et *Homo* (première occurrence vers environ 2,8 Ma) qui donneront elles-mêmes plusieurs espèces (Behrensmeyer and Reed, 2013; Campisano et al., 2017). Le rift est africain restera un centre de diversité et de diversification majeur pour les Hominini jusqu'à environ 1,4 Ma (Maslin et al., 2015). C'est également dans cette région d'Afrique que sont observés les premières traces d'utilisation d'outils visibles sur des os à 3,39 Ma (Dikika, base vallée de l'Awash, Ethiopie) (McPherron et al., 2010) suivi des premiers outils en pierre à 3,3 Ma (Lomekwi, berge ouest du lac Turkana, Kenya) (Harmand et al., 2015). Cette branche du rift constitue entre 4,3 Ma et 1,4 Ma une zone spatiale et temporelle charnière dans l'histoire de l'évolution de la lignée humaine, aussi bien au niveau morphologique que culturel. C'est également un lieu important pour comprendre l'apparition de notre genre, le genre *Homo* qui trouverait ses origines dans la base vallée de l'Awash (Ethiopie) vers 2,8 Ma avec une hémimandibule gauche (LD350-1) identifiée comme *Homo* sp. (Villmoare et al., 2015).

---

<sup>a</sup> Dans ce manuscrit, nous utiliserons la classification taxonomique la plus récente où le terme Hominini réfère à *Homo sapiens* et aux espèces disparues plus étroitement liées à *Homo sapiens* qu'au genre *Pan* (van Holstein and Foley, 2017).



**Figure 1.1** - Répartition chronologique et spatiale des espèces d'Hominini adaptée d'après Maslin et al. (2015) et complétée avec Bobe et al. (2020), Behrensmeyer et Reed (2013), Campisano et al. (2017) et Levin (2015). Elle se base sur les dates de premières et dernières occurrences.

Plusieurs théories expliquent l'évolution des Hominini (Maslin et al., 2015). Celles-ci admettent que les variations environnementales aient pu jouer un rôle prépondérant dans la spéciation entre 4,3 et 1,4 Ma. Or, des changements de végétations importants, comme l'expansion des graminées en C<sub>4</sub>, sont survenus en Afrique de l'Est pendant cette période, mais ceux-ci ne sont pas synchrones et varient en fonction des lieux (Levin, 2015). La répartition des paléoenvironnements varie considérablement, tant dans l'espace que dans le temps, les relations entre l'évolution des Hominini et ces tendances sont donc discutables. Une meilleure compréhension des habitats occupés par les Hominini est donc nécessaire pour appréhender le rôle et l'importance des différents mécanismes évolutifs dans l'évolution de la lignée humaine, que ce soit pour le développement de certains traits comme la bipédie ou le développement des industries lithiques.

Bien qu'intéressants et informatifs, les enregistrements suprarégionaux comme ceux provenant des archives marines (Bonnefille, 2010) ne permettent pas de comprendre comment les changements de végétations ont eu lieu localement, c'est-à-dire au niveau des habitats. Pour mieux comprendre comment l'évolution de la lignée humaine a pu être influencée par des changements d'habitats, des enregistrements continentaux qui rendent compte de la diversité spatiale de la végétation à un temps donné sont donc nécessaires. Or à ce jour ils sont peu nombreux (Bonnefille et al., 1987; Bonnefille et al., 2004; de Heinzelin, 1983; López-Sáez and Domínguez-Rodrigo, 2009; Saylor et al., 2019), et indiquent souvent que les Hominini fossiles auraient été associés à des environnements dits mosaïques (Reynolds et al., 2015). A l'instar des conclusions obtenues à partir des analyses paléontologiques (Behrensmeyer and Reed, 2013), nous pensons que les environnements mosaïques pourraient être des interprétations erronées liées à une résolution spatiale trop faible des reconstructions paléoenvironnementales.(Bonnefille et al., 2004)

Dans le but d'améliorer la résolution spatiale des reconstructions paléoenvironnementales et la compréhension des préférences écologiques des Hominini, nous présentons ici quatre études. Celles-ci se présentent sous forme d'articles scientifiques (rédigés en anglais), dont un est publié (Bourel et al., 2020). Ces travaux s'inscrivent dans la continuité des études polliniques menées dans le nord du rift est africain par l'équipe de Raymonde Bonnefille à partir des années 1970 et jusque dans les années 2000 qui visaient à documenter les environnements plio-pléistocènes (Bonnefille, 2010). Pour ces quatre études, l'accent a été mis sur la reconnaissance des micro-habitats. Chaque article se lit indépendamment des autres, et présente la zone d'étude et un état de l'art du sujet abordé. Nous ne reprendrons donc pas ces points ici.

## 1.2. Présentation des études

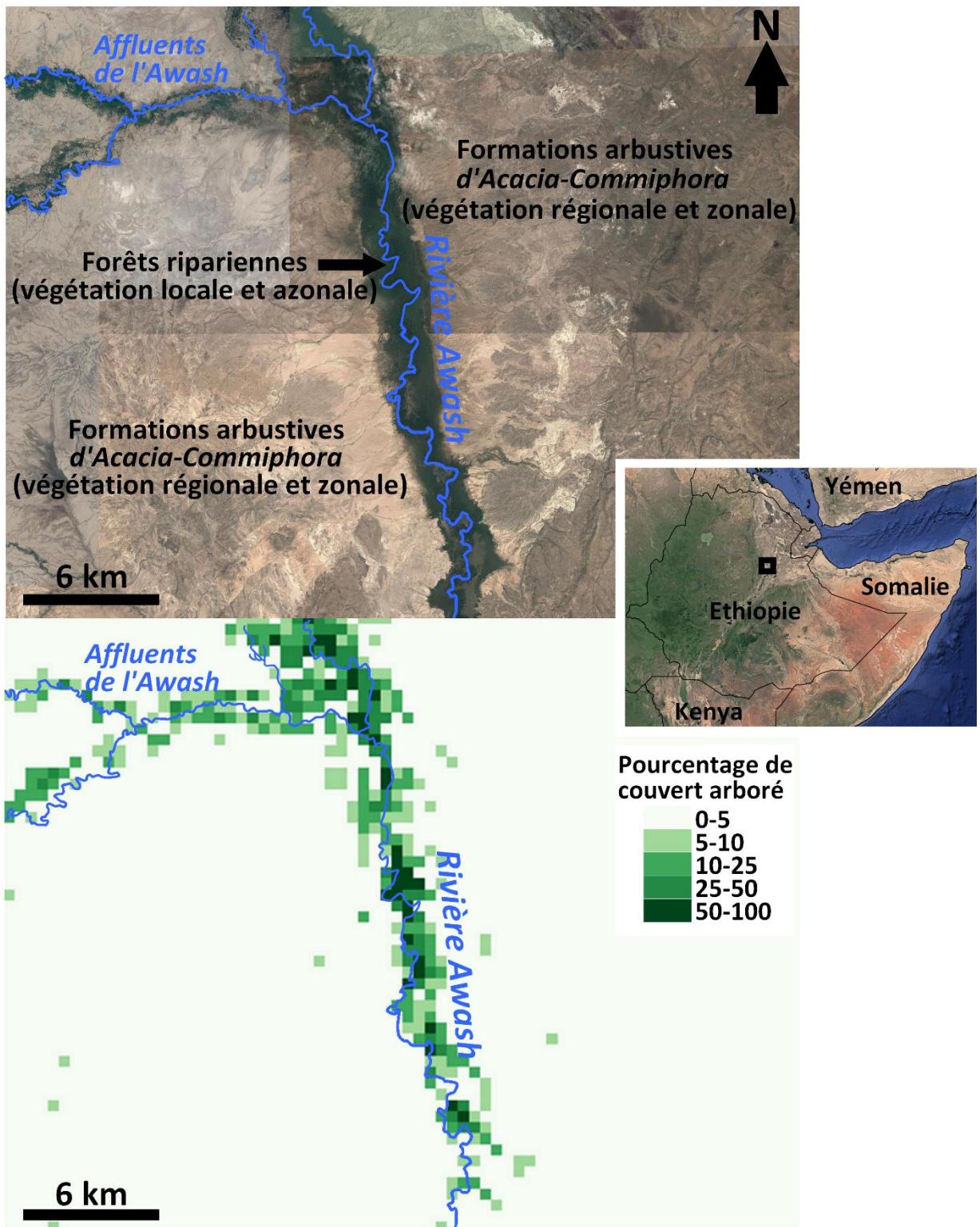
**La première étude est intitulée « Inférence des types de végétations locales et régionales à partir des assemblages polliniques : une nouvelle approche pour étudier la paléoécologie des premiers Hominini »** (page 21). Elle propose une nouvelle approche de calcul de scores polliniques avec des intervalles de significativité qui

reflètent l'affinité d'un spectre pollinique pour différentes végétations. Elle permet aussi d'extrapoler des variables climatiques et des pourcentages de couvert arboré.

C'est une nouvelle méthode de calibration du proxy pollen, développée en vue d'une utilisation dans le registre fossile. Avant nos travaux, seule la méthode des *Plant Functional Types* (PFT) ou biomisation (Prentice et al., 1992), permettait une interprétation objective des données polliniques en termes de paléovégétation. Elle a été appliquée aux données Pliocène de l'Afar en parallèle à la méthode des *Best Analog* (BA) qui a fourni une quantification du paléoclimat (Bonnefille et al., 2004). La méthode des PFT permet d'identifier un biome, c'est à dire un type de végétation reconnaissable à l'échelle suprarégionale, qui est en équilibre avec le climat. Elle néglige le signal des végétations ripariennes dont la présence est liée à des facteurs édaphiques locaux (Bonnefille et al., 2004; Jolly et al., 1998; Peyron et al., 2000; Prentice et al., 1992). Or, on sait aujourd'hui que les sites à Hominini sont très fréquemment associés à des contextes ripariens ou de résurgences d'eaux souterraines (Barboni et al., 2019; Campisano and Feibel, 2008; Levin, 2015). De plus, les botanistes ont montré que les végétations ripariennes en Ethiopie sont représentées par de nombreux sous-types de végétations dont la composition floristique diffère (surtout en dessous de 1800 m) de la végétation environnante avec laquelle elles partagent plus ou moins des espèces similaires (Friis et al., 2010a). Ces végétations ripariennes sont des végétations locales et sont dites azonales car elles ne sont pas en équilibre avec le climat. Dans les régions très arides, telle que l'Afar, les formations ripariennes azonales sont également caractérisées par un couvert arboré largement supérieur à celui des steppes et savanes environnantes (**Figure 1.2**). Ces environnements peuvent alors à tort être interprétés comme des environnements mosaïques si la résolution spatiale des reconstitutions est insuffisante. En palynologie, il est difficile de différencier le signal pollinique local du signal régional, que ce soit par la méthode des PFT ou par l'approche classique d'interprétation des données basée sur la connaissance de l'écologie des espèces représentées par les taxons polliniques (Bonnefille et al., 1987). L'objectif de la nouvelle approche développée a donc été de séparer, à partir des spectres polliniques, les sous-types de végétations que l'on observe à l'intérieur des végétations régionales et plus particulièrement les végétations locales liées à un contexte riparien ou de résurgence.

Cette approche a été construite et calibrée à partir d'un jeu de données polliniques modernes et de données satellitaires sur la période 1970 à 2000 pour le climat, 2000 à 2004 pour la production primaire nette, et de 2008 à 2017 pour le couvert arboré et les autres données de *Land Cover*. Le jeu de données polliniques modernes obtenu à partir de 283 d'échantillons de surface inclut 62 échantillons nouveaux (soit 22% du jeu de données). Ces 62 nouveaux échantillons proviennent en majorité d'une campagne de prélèvement réalisée en septembre 2017 dans le cadre de cette thèse. Ces échantillons complètent des manques de données polliniques pour les végétations des formations arbustives à *Acacia-Commiphora* avec des précipitations moyennes annuelles (PANN) inférieures à 400-600 mm/an et des températures moyennes annuelles (TANN) comprises entre 20-30°C, et pour les végétations de zones afromontagnardes avec des PANN entre 700-800 mm/an et des TANN entre 18-21°C. Afin d'obtenir les meilleurs résultats possibles, la géolocalisation de certains anciens spectres polliniques a été retravaillée à partir de nombreux documents (eg., notes d'auteurs, cartes, etc). Sans ce travail qui a permis de corriger les positions de 68% des anciens spectres (150/221 spectres) avec une correction moyenne de 9,7 km, il aurait été impossible de comparer les données polliniques avec les nombreuses données climatiques et environnementales des satellites via un système d'information géographique (SIG). Dans la même optique d'améliorer la qualité des données, les noms des types polliniques ont été unifiés entre tous les échantillons et une liste des types synonymes a été développée en s'appuyant sur le travail de Vincens et al. (2007).

**La seconde étude est intitulée «*Australopithecus afarensis* relié aux environnements boisés : reconstruction palynologique à haute résolution spatiale pour les paléo-végétations des Formations de Hadar et Busidima (Afar, Ethiopie) »** (page 63). En continuité directe avec la précédente, cette étude applique notre nouvelle approche de calculs de scores polliniques à des spectres polliniques fossiles provenant des Formations pliocènes de Hadar et Busidima dans la basse vallée de l'Awash (Ethiopie). Ces formations sont riches en fossiles d'*Australopithecus afarensis* dont la présence semble avoir été continue dans cette région entre 3,42 Ma et 2,96 Ma (Campisano, 2007; Reed, 2008). Nous nous sommes intéressés aux habitats d'*Australopithecus afarensis* via une ré-analyse des données polliniques entre 4 Ma et 2,35 Ma afin de couvrir l'ensemble de la période d'occurrence *Australopithecus afarensis* (entre 3,42 Ma et 2,96 Ma) dans la basse vallée de l'Awash.



**Figure 1.2** - Développement local de forêts azonales le long de la rivière dans une végétation régionale et zonale de formations arbustives à *Acacia-Commiphora* (moyenne vallée de l'Awash, Ethiopie). Pourcentage de couvert arboré issu de Global Maps (année 2008,  $0.46 \times 0.46$  km, Kobayashi et al., 2017)

Il nous est apparu important de réétudier les spectres polliniques de ces formations car malgré la présence relativement continue d'*Australopithecus afarensis* entre 3,42 Ma et 2,96 Ma, il n'a pas pu être dégagé de régularité à l'échelle des biomes ou de la végétation régionale pour expliquer le maintien de cette espèce (Bonnefille et al., 2004). Or durant cette période, *Australopithecus afarensis* aurait cohabité dans la basse vallée de l'Awash (Ethiopie) avec au moins deux autres espèces d'*Australopithecus* : une forme transitionnelle d'*Australopithecus anamensis* de 3,8 Ma à 3,6 Ma (Haile-Selassie et al., 2010b) et *Australopithecus deyiremeda* vers 3,3-3,5 Ma (Campisano et al., 2017). Ces différentes espèces occupaient-elles différents habitats que l'analyse palynologique classique n'a pas été en mesure d'identifier ? Dans la mesure où il est difficile de concevoir une telle diversité d'espèces d'Hominini sans la présence d'habitats ou de préférences écologiques distinctes (Behrensmeyer and Reed, 2013), il est apparu tout à fait justifié d'appliquer notre nouvelle approche qui prend en compte la paléovégétation non plus seulement à l'échelle régionale, mais aussi à l'échelle locale. Cette étude vise donc à voir si *Australopithecus afarensis* est bien une espèce ubiquiste comme les reconstitutions environnementales à l'échelle régionale le proposent (Bonnefille et al., 2004; Reed et al., 2013), ou une espèce avec une préférence écologique.

Pour ce faire, nous avons rassemblé les données de 33 spectres polliniques anciens (Bonnefille et al., 1987; Bonnefille, 2010; Bonnefille et al., 2004; López-Sáez and Domínguez-Rodrigo, 2009; Saylor et al., 2019) et acquis 9 nouveaux spectres du grand sondage terrestre NAW-1A (basse vallée de l'Awash, Ethiopie) issus du programme international HSPDP (Cohen et al., 2016). Le grand nombre d'échantillons stériles (144) sur les 175 échantillons des sondages HSPDP NAW-1A et NAO-1B (à 3 km de NAW-1A) que j'ai préparé n'a pas permis d'augmenter significativement la résolution temporelle. Néanmoins, les 9 nouveaux spectres polliniques obtenus ont permis de combler un manque de données entre 3,43 et 3,39 Ma.

Étant donné que nous devions travailler à partir des données polliniques brutes, les anciens spectres polliniques ont dû être saisis pour la plupart (27/33) à partir des archives de l'équipe de palynologie du CEREGE. Les types polliniques ont été harmonisés à partir de la liste des synonymes mise en place pour la première étude sur les données modernes. Un important travail bibliographique et d'étude des rapports de missions archivés au laboratoire de palynologie du CEREGE a aussi été fait pour replacer

l'ensemble de ces échantillons dans la nouvelle stratigraphie composite de la Formation de Hadar et Busidima (Campisano, 2007; Campisano and Feibel, 2008).

**La troisième étude est intitulée « Enregistrement pollinique et dynamique des microhabitats associés à la résurgence d'eau douce souterraine de Kisima Ngeda, lac Eyasi, Tanzanie »** (page 93). Cette étude présente l'analyse pollinique d'un court sondage couvrant les derniers 1250 ans (entre 760 AD et 2011 AD) dans le but de mieux comprendre la dynamique de la végétation liée à des sources d'eaux souterraines, qui sont des hydrosystèmes clés pour les faunes en zones arides (Barboni et al., 2019). Ces hydrosystèmes sont fréquents dans le rift non seulement aujourd'hui (Cuthbert et al., 2017), mais aussi dans le passé. Notamment au niveau des sites à Hominini d'Olduvai vers 1,8 Ma (eg., Ashley et al., 2009; Barboni et al., 2010), des sites pliocènes dans la moyenne vallée de l'Awash (WoldeGabriel et al., 2009) et dans d'autres partie du rift est africain (voir synthèse dans Barboni et al. 2019). Ainsi qu'à Hadar aux niveaux des sites à *Australopithecus afarensis* entre  $3.372 \pm 0.01$  et  $3.370 \pm 0.01$  Ma, comme l'a mis en évidence notre travail précédent (**Partie 3 de cette thèse**).

L'analyse pollinique de ce court sondage a été menée dans le but de savoir comment ces sources influencent les mouvements et/ou remplacements de végétations au sein des habitats mosaïques et à quelle vitesse. Dans le site d'Olduvai, situé à moins de 50 km du Lac Eyasi, il est possible d'observer via les travertins des durées d'occurrences de système de sources d'environ 20 000 ans qui sont interprétés comme des décharges d'eaux souterraines continues (Ashley et al., 2009). Mais, ces données géologiques n'ont pas une résolution suffisante pour déterminer si le débit des eaux souterraines de ce type de résurgences était stable ou épisodique pendant ces longues périodes d'activation supposée. L'hypothèse d'une décharge ponctuelle des eaux souterraines liée à des variations climatiques décennales ou centenaires ne peut être exclue car non observable dans le registre fossile ou géologique faute d'une résolution temporelle suffisante. Réfuter cette hypothèse permettrait d'appuyer l'idée que ce type de système peut soutenir des végétations locales et boisées de façon stable malgré des épisodes climatiques secs pouvant s'étendre sur plusieurs décennies ou siècles. Cela permettrait d'appuyer l'importance de ces micro-habitats pour les populations d'Hominini.

**Enfin, la quatrième et dernière étude présentée dans ce mémoire de thèse est intitulée « Reconnaissance automatisée de grains de pollens modernes, fossiles, intacts et endommagés par réseaux neuronaux convolutifs multiples »** (page 121). Cette étude, publiée (Bourel et al., 2020), s'est intéressée à la capacité des systèmes de reconnaissances automatiques à assister les palynologues dans la détermination et le comptage des grains de pollens en contexte paléontologique. Elle vient en perspective de futurs travaux dans la continuité des études présentées ci-dessus . En effet, durant ce doctorat, il s'est avéré que la réalisation des comptages polliniques au microscope était particulièrement difficile et chronophage, en particulier dans les niveaux du Pliocène des sondages NAW-1A et NAO-1B (basse vallée de l'Awash, Ethiopie). Ceci est dû à la pauvreté des sédiments en grains de pollens mais aussi au mauvais état de conservation de ces derniers. Au cours de cette thèse, sur les 175 échantillons des carottes NAW14-1A et NAO14-1B du projet HSPDP que j'ai traités par digestion acide au laboratoire, 144 échantillons (82%) se sont révélés stériles ou sub-stériles, 22 échantillons (13%) présentaient des pollens en très faible quantité avec une mauvaise préservation et n'étaient pas comptables manuellement. Au final, seuls 9 échantillons (5%) ont pu être comptés avec un temps de comptage moyen d'une semaine par échantillon. Ce type de problème n'est malheureusement pas limité aux sites que nous avons étudiés. Il est courant à l'ensemble des sites à Hominini du Plio-Pléistocène du rift est-africain (Bonnefille, 1984; de Heinzelin, 1983). Au vu des résultats sur d'autres types de microfossiles (Beaufort and Dollfus, 2004), l'utilisation d'un système de comptage automatique permettrait probablement de fortement réduire le temps des comptages et d'analyser des échantillons qui sont actuellement délaissés car beaucoup trop chronophage. De nombreuses approches ont été développées pour automatiser les comptages polliniques notamment en aéropalynologie, mais toutes ces approches ne fonctionnent que sur des grains frais et intacts (Holt et al., 2011; Holt and Bennett, 2014). Des tests de reconnaissance automatique ont été concluants dans des tourbières mais toujours sur des pollens très bien préservés (Han and Xie, 2018). Le développement d'un système capable de reconnaître automatiquement des pollens endommagés serait donc plus approprié. Dans ce but, il est présenté ici les résultats d'un système basé sur des réseaux de neurones artificiels de type *convolutional neural networks* (CNNs) qui a été testé sur des pollens présentant différents degrés de préservation.

### **1.3. Développements et valorisations annexes**

Pour terminer, les données et analyses réalisées durant cette thèse ont également contribué à d'autres études qui ont fait l'objet de deux publications scientifiques auxquelles j'ai été associé comme co-auteur, une dans la revue *Review of Palynology and Paleobotany* (Barboni et al., 2019) et une dans la revue *Nature* (Saylor et al., 2019). Ces publications ne sont pas présentées ici en détail mais sont visibles en **Annexe paper 1.1** et **Annexe paper 1.2**. La première de ces publications (Barboni et al., 2019) porte sur l'importance de prendre en compte les forêts et environnements boisés associés aux milieux humides édaphiques dans l'interprétation paléoécologique des sites à Hominini. Elle intègre des échantillons polliniques de sols modernes prélevés au niveau de résurgences d'eaux souterraines que j'ai échantillonnés lors d'une campagne de prélèvement réalisée en 2017 en Ethiopie dans le cadre de cette thèse (5/41 échantillons). Mon implication dans cet article, en plus d'avoir recherché et échantilloné les sites, a principalement concerné la préparation des échantillons polliniques, le comptage des pollens et l'interprétation des spectres polliniques. Le travail effectué au niveau des comptages polliniques a ensuite été réutilisé pour la première étude de ce manuscrit (**Partie 2 de cette thèse**). Le second article (Saylor et al., 2019) a porté sur l'analyse des paléoenvirons, des conditions de dépôts et des âges des sédiments de la vallée de Godaya (Woranso-Mille, Ethiopie) dans lesquels a été retrouvé un crâne d'*Australopithecus anamensis* particulièrement bien préservé (spécimen MRD-VP-1/1) décrit dans Haile-Selassie et al. (2019). On sait désormais que ce crâne âgé de 3,8 Ma revêt une importance particulière pour la compréhension des relations taxonomiques et phylogéniques entre *Australopithecus anamensis* and *Australopithecus afarensis*. Mon implication dans ce papier a d'abord été la préparation et le comptage de 16 échantillons polliniques provenant du site d'étude du spécimen MRD-VP-1/1 (vallée de Godaya, Woranso-Mille, Ethiopie). Ces échantillons n'étaient malheureusement pas exploitables pour des analyses polliniques (stériles ou trop pauvres en pollens) à l'exception l'échantillon MRD-210 provenant du même niveau que le crâne de MRD-VP-1/1. Dans cet article de Saylor et al. (2019), j'ai aussi proposé une reconstitution de l'habitat du spécimen MRD-VP-1/1.

## **2. Inférence des types de végétation locale et régionale à partir des assemblages polliniques : une nouvelle approche pour étudier la paléoécologie des premiers Hominini**

- Titre de l'article : Inferring local and regional vegetation types from pollen assemblages: a new approach for investigating early hominin paleoecology.

- Auteurs : Benjamin Bourel<sup>1</sup>, Doris Barboni<sup>1</sup>, Jean-Charles Mazur<sup>1</sup>

<sup>1</sup>CEREGE, Aix Marseille Université, CNRS, IRD, INRAE, Coll. France, Technopole Arbois, 13545 Aix en Provence cedex 4, France

- Revue : Sera soumis à *Journal of Biogeography* pour leur section *Methods and Tools*

- Résumé : Hominini sites in the Plio-Pleistocene of Eastern Africa have long been known to be associated with riparian environments. Currently, in the rift, these riparian environments can allow the development of a wide variety of local vegetation that differs greatly from the surrounding regional vegetation. Nevertheless, due to the paradigm that Hominini speciation is linked to large-scale vegetation changes and the difficulty of differentiating between local and regional pollen signals, palynological research has focused on the study of vegetation variations at regional scale. Today, this paradigm is being questioned and descriptions of the local riparian vegetation of paleoanthropological sites remain vague. In order to pursue palynological studies at the regional scale but also at the local scale at Hominini sites, we propose in this study a new approach to calculate pollen scores. Our new approach is intended to be objective and it is based on a statistical and machine learning analysis on a modern data set (mainly Ethiopian) of 283 surface samples (muds and soils). This approach allows the identification of 24 types and sub-types of vegetation, nine of which are local riparian and azonal vegetation. It differs from previous pollen approaches in the spatial precision of the reconstruction of vegetation (lower than the regional scale), the integration of riparian vegetation and, further treatment of aquatic plant pollen. With our approach, we highlight the importance of treating pollen data differently depending on the conditions of deposit : abundance classes for non-dynamic deposition environments (e.g. soils or lake edges) and presence-absence for dynamic deposition environments (e.g. river beds or edges). We show that the interest of this approach is not limited to the determination of paleovegetation but

also to the inference of classically reconstructed environmental parameters (precipitation, temperature, etc.) and new very interesting parameters such as the percentage of forest cover, rainfall regime or river morphology.

- Mots clefs : Ethiopia, machine learning, spring, riparian vegetation, ecology

## 2.1. Introduction

Many paleontological areas in Ethiopia have preserved abundant fauna, including hominin remains of Late Miocene, Pliocene, and Pleistocene age (eg., Campisano et al., 2017; Haile-Selassie et al., 2015; Johanson, 2017; Saylor et al., 2019; Villmoare et al., 2015). Ethiopia is an important place to study the history of human evolution and to understand how the environment and climate may have influenced it (Levin, 2015). For some paleontological areas in Ethiopia, such as the *Australopithecus afarensis*-bearing Hadar Formation (Bonnefille et al., 1987), the *Ardipithecus ramidus*-bearing Aramis member (WoldeGabriel et al., 2009), more recently the *Australopithecus anamensis*-bearing deposits at Woranso-Mille (MRD locality) (Saylor et al., 2019) and other *Australopithecus* sites such as Meshellu/Dikika also in the North Afar (Bonnefille, 2010), pollen were extensively studied and provided direct botanical evidences of the plant taxa associated with early hominids. By comparison with modern pollen data obtained from surface soil samples, fossil pollen abundances allowed to appraise the composition of past plant communities (Bonnefille et al., 1987). Fossil pollen assemblages (from Hadar only) were also interpreted in terms of Plant Functional Types (PFTs) and PFT associations or plant biomes using the biomisation procedure (Prentice et al., 1992). Constrained by the climatic requirements of PFTs and biomes, the fossil pollen data of Hadar provided quantitative estimates of the paleo-climate between 3.4-3.18 million years (Ma) (Bonnefille et al., 2004). Plant species occurrence and abundance today is the result of both climatic and non-climatic factors (eg., species-specific pests). Hence, paleo-climate estimates and biomes inferred from PFTs climatic requirements rather than from the present-day distribution of plant species are likely robust (Harrison et al., 2010).

Applied to Hadar, the biomisation method initially calibrated by Jolly et al. (1998) for Africa and the Arabian Peninsula, and subsequently modified by Peyron et al. (2000) and Bonnefille et al. (2004) provided paleo-temperature, precipitation, and aridity

estimates showing cooling and reduced aridity around 3.3 Ma within the 3.4-3.18 Ma interval. Yet, the reconstructed paleovegetation remain dominated by biomes such as tropical xerophytic wood/scrub, temperate xerophytic wood/scrub, and steppe, which got very similar score values throughout the sequence. Reconstructing biomes *in equilibrium* with climate may provide plausible paleo-climate estimates, but is a too coarse approach to capture the habitat diversity that likely made the Pliocene North Afar an attractive place for a wide variety of animals (terrestrial, aquatic, grazers and browsers, frugivorous primates including several hominin species) as recorded by the abundance and diversity of faunal taxa in this area (Faith et al., 2019; Reed et al., 2013). Although the presence of regional vegetation types and sub-types with different climatic conditions is clearly established by botanists in this part of the Rift (Friis et al., 2010a), the potential of pollen to characterize them objectively has not been developed.

In arid and semi-arid eastern Africa, surface water such as rivers, lakes and groundwater discharge areas (springs) promotes the development of micro-habitats, and strongly influences vegetation and wildlife distribution (eg., Friis et al., 2010; Geheb and Abebe, 2003). In arid and semi-arid Afar region for example, locally wooded vegetation can be observed within the otherwise grass- or dwarf-shrub dominated vegetation that occurs at the landscape-scale, such as a dense riparian forest along the Awash river and *Hyphaene* palm woodlands near groundwater discharge areas (springs) (Friis et al., 2010a; Geheb and Abebe, 2003). These azonal wooded riparian vegetation associated with permanent water were particularly attractive for hominids in the past (Barboni et al., 2019). Yet, few vegetation proxies allow reconstructing paleovegetation at the spatial scale of hominid sites. Phytoliths and organic compounds were used for this purpose at a few sites in Ethiopia (WoldeGabriel et al., 2009) and Tanzania (eg., Arráiz et al., 2017; Magill et al., 2016), but their taxonomic resolution is too low to characterize paleovegetation beyond its physiognomy (ie. the paleo-tree cover), or beyond indicating the presence of palms, sedges, ferns, or details within the grass paleo-communities. To our knowledge, there is no more accurate proxy than pollen to describe the composition of past vegetation. Yet, there has been no interest, so far, to give importance to the local vegetation (riverine and fed by springs) signals contained in pollen assemblages despite the fact that “riverine forests are represented by a long range of floristic sub-types from nearly all parts of Ethiopia”, and differ in composition according to topographic and climatic factors, and the surrounding vegetation with which they more or less share

similar species (Friis et al., 2010a, p. 36). The potential of pollen proxy to characterize different riparian vegetation types, which typically occur as narrow and strip-like habitats along perennial and non-perennial rivers below 1800 m above mean sea level (a.s.l.) and associated with lakes and springs, has yet to be formally tested.

To test the potential of pollen assemblages to infer regional (climatically driven) and local (edaphically-driven) vegetation types at hominid sites in Northeastern Africa, it is necessary to use a well-documented modern pollen dataset. Initiated by Bonnefille in the 1970s for Eastern Africa, surface soils sampling has provided modern benchmarks against which fossil pollen data could be compared. This uniformitarian approach is possible because in Eastern Africa floristic composition of plants ecosystems is relatively stable since the last 11 Ma, despite multiple and reversible changes of geographical expansions of plants ecosystems (Bonnefille, 2010). To date, the African Pollen Database includes 128 modern samples from Ethiopia and the Arabian Peninsula (“African Pollen Database,” 2019), but 83 additional modern samples from published studies (Bonnefille et al., 1987, 1993; Bonnefille and Buchet, 1986; Mohammed, 1992), including five from spring-fed wetlands and forests (Barboni et al., 2019) can be added. However, this dataset still lacks of pollen samples from low elevation, notably from the Awash Valley and from the southern extremity of the Ethiopian Plateau, where rainfall is bimodal.

Pollen data are more often expressed in terms of percentages of taxa than in terms of presence/absence as it is generally the case with paleontological (faunal) data. However, as one can observe studying modern pollen assemblages, the abundance approach is prone to giving a lot of discriminating power even with small differences of percentages, which can be a serious issue when dealing with fossil assemblages that definitely underwent taphonomical processes. On the other hand, ignoring the difference of pollen percentages resulting from a real difference in the vegetation between two sites is also not satisfactory. The use of an abundance scale is a compromise that still needs to be tested.

Hence, because it is difficult to conceive a diversity of *Australopithecus* species without a diversity of environments (Behrensmeyer and Reed, 2013), we sought to develop a new approach based on the pollen proxy that would allow inferring regional and local paleovegetation within the biomes of Eastern Africa. In this study we present

the results of a calibration study we have carried out on a newly compiled and implemented modern pollen dataset for Ethiopia, which includes new lowlands samples from the Awash Valley and the southern extremity of the Eastern Highlands that complement existing data. This new approach uses pollen data expressed in terms of abundance classes defined by statistical analyses. We have developed a script in R, which ensures that the interpretation of the pollen signal in terms of vegetation is objective and standardized. This script applies hierarchical and *k*-means clustering to identify which modern vegetation types can be discriminated using pollen scaled-abundance. It then uses Value-Test to measure the affinity of every pollen taxon for every vegetation type and sub-type. These Value-Tests of pollen taxa for vegetation types and sub-types can then be used to evaluate how a pollen assemblage (modern or fossil) relates to the different regional and local vegetation types present today in Ethiopia.

## 2.2. State of the art

The development and use of approaches involving the calculation of pollen scores for fossil sites in Africa is relatively old. The first approach that has been implemented is the Best Analogues (BA) method. La BA method was first developed for European pollen data in 1990 (Guiot, 1990) and was then applied for the first time in Africa by Bonnefille et al., 1992 on Kashiru holocene samples (Burundi). The BA method uses statistical analysis centered on a chord distance to only reconstruct climatic parameters from fossil pollen spectra. It is a robust method that has been used in Western, Central and Eastern Africa without significant modification (Bonnefille et al., 2004; Vincens et al., 2006a). The second and last approach used in Africa is the plant functional types (PFT) method. Based on the work of Prentice et al. (1992) on the characterisation of biomes in Europe by a combination of PFTs, Jolly et al. (1998) proposed a pollen biomisation method for Africa and the Arabian Peninsula based on the allocation of one or more PFTs to the different pollen types and calibrated with 966 modern pollen samples from Africa, sub-Saharan Africa and the Arabian Peninsula. The PFT method are based on the known modern ecological range of pollen-producing plants and allows to reconstruct both biomes and climatic parameters. Jolly et al. (1998) thus produce a PFT method to reconstruct 11 biomes. The results are conclusive for nine biomes, but the desert and dry tropical forest biomes remain perfectible (Vincens et al., 2006a).

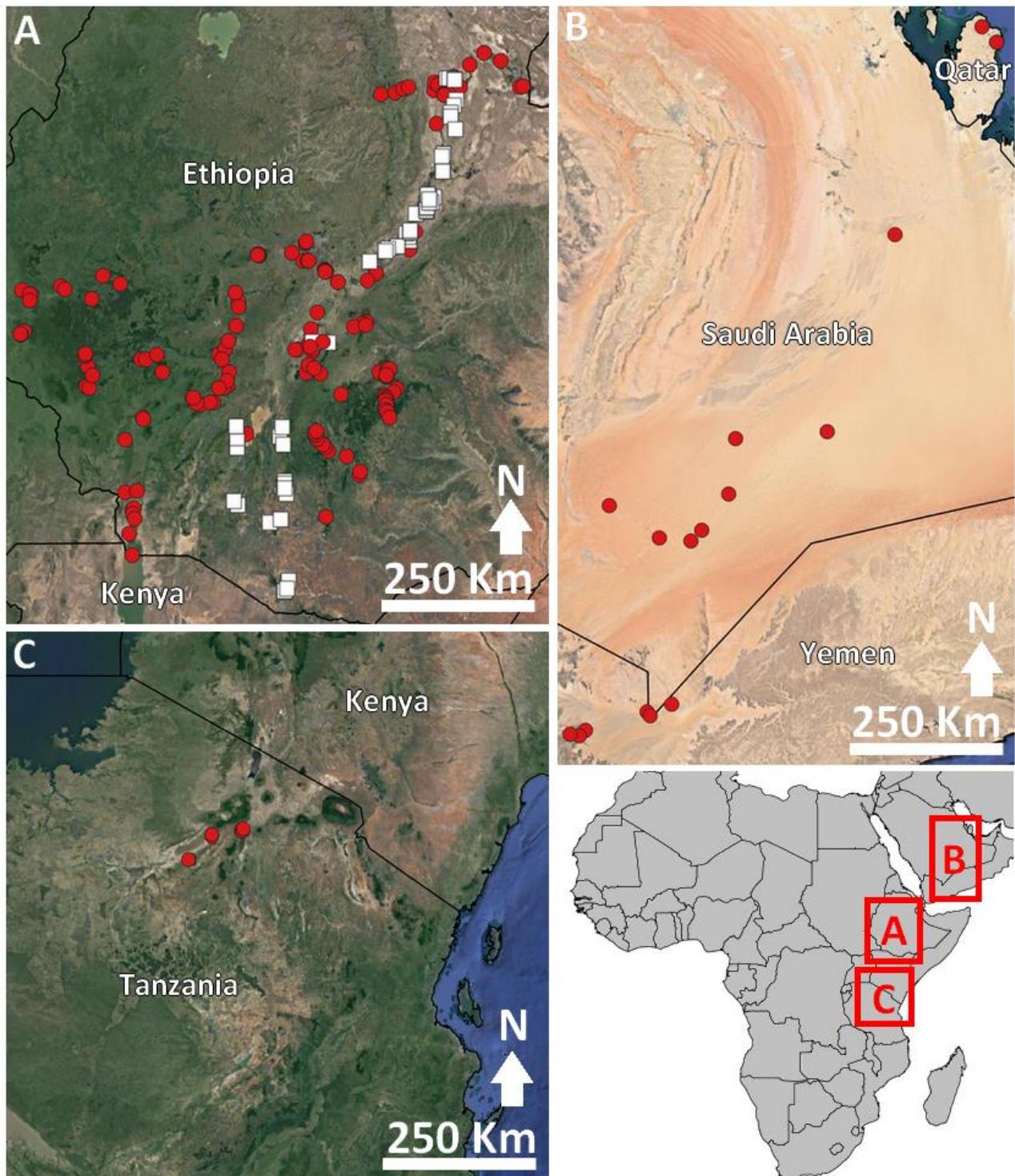
The PFT method of Jolly et al. (1998) has reused by Peyron et al. (2000) who grouped tropical evergreen, wet tropical evergreen, and dry tropical evergreen PFTs in tropical evergreen PFT, added two new PFTs (grass and ubiquitous) and calibrated this version of PFT method with 227 modern pollen samples of Ethiopia, Kenya, Uganda, and Tanzania compiled by Bonnefille et al. (1992). Peyron et al. (2000) produced a spatially consistent reconstruction of climate parameters at present and at 6000 14C yr B.P. for 32 sites in East Africa. However, the interpretation by the TFP method of past and present vegetation is unfortunately not very clear. The biomes for the present are not very informative with a majority of sites defined as “tropical dry forest or savanna” for example. Bonnefille et al. (2004) also reused the method of Jolly et al (1998) for the reconstruction between 3.4 Ma and 2.9 Ma at the level of the Hadar Formation. This time, the same modern soil samples as Jolly et al. (1998) are used for the calibration. But only four biomes are used. Three biomes of Jolly et al. (1998) (tropical xerophytic woods/scrub, temperate xerophytic woods/scrub, and the steppe biomes) and a new alternative biome of warm mixed forest which is the fusion of the tropical rain forest, seasonal forest, tropical dry forest, and warm mixed forest biomes of Jolly et al. (1998). The savanna and desert biomes of Jolly et al. (1998) were not used by Bonnefille et al. (2004). Bonnefille et al. (2004) have very consistent results for paleoclimatic parameters. However, the results of the PFT method are very complex to interpret because the biome signal is dominated by tropical xerophytic woods/brushes, temperate xerophytic woods/brushes and steppe biomes. These three biomes have similar score values throughout the sediment record.

In 2006, Vincens et al. (2006) proposed a new modified version of Jolly et al. (1998)'s method with the grouping of tropical evergreen, wet tropical evergreen, and dry tropical evergreen PFTs in tropical evergreen PFT and the addition of grass PFT. The nomenclature of pollen types has also modified to match with the nomenclature of African Pollen Database and the PFTs assignments for the pollen types are also slightly modified . They worked on 150 modern samples of south of Eastern Africa (Kenya, Tanzania, Uganda). Their result is correct for 82.6% of sites for broad-definition biomes (e.g. savanna, steppe). Vincens et al. (2006) highlight the difficulties of reconstructing the mosaic of open and closed vegetation, which is the most frequently defined habitat type for early hominids (Reynolds et al., 2015; Stewart, 2014). The development of Jolly et al. (1998)'s method continued for central and Western Africa (Hély et al., 2006; Izumi and

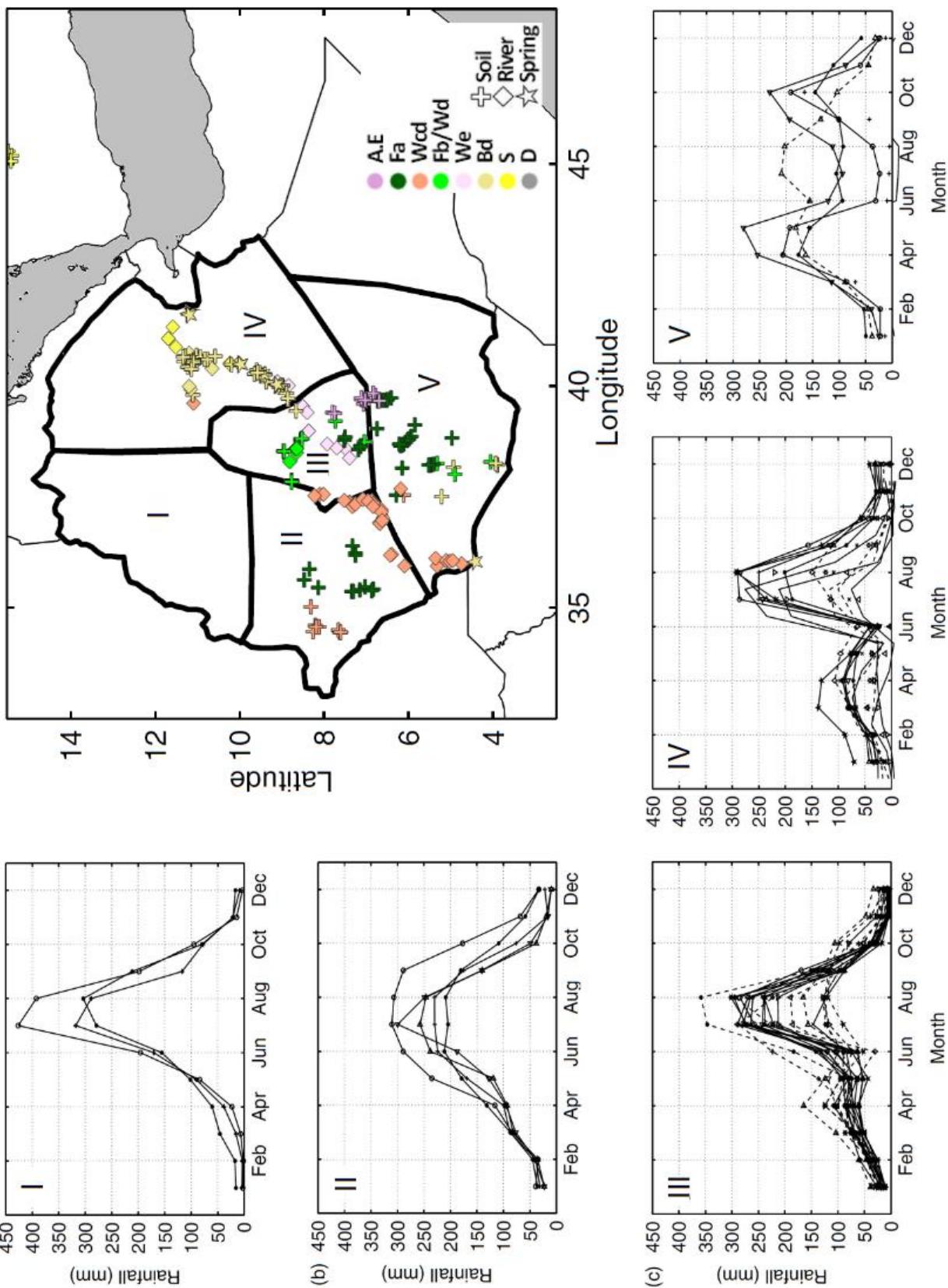
Lézine, 2016; Lebamba et al., 2009) but did not evolve significantly for Eastern Africa after Vincens et al. (2006). In 2016, Izumi and Lézine used biomisation on 130 modern samples from East Africa. Their biomisation is very close to that of Vincent with some variations such as the limitation of the Poaceae to the savanna, steppe and desert biomes. Their best result is lower than that of Vincens et al. (2006) with a correct classification of 74.6%. The main problems of biomization concern the sites of the savanna and steppe biomes with respectively only 50% and 69% of successful classification.

### 2.3. Study area

This study focuses on the present-day vegetation and climate of the Horn of Africa, specifically in Ethiopia, for which pollen assemblages from surface soil samples are available (**Figure 2.1**). In Ethiopia, topography, which is with the seasonal movement of the Intertropical Convergence Zone (ITCZ) one of the main factors influencing precipitation and precipitation patterns, is characterized by highlands generally between 1600 and 2600 m a.s.l. (up to 4500 m a.s.l.) of elevation on the western and eastern sides of the Rift valley. In Central Ethiopia, the floor of the Rift valley, where large lakes are located, lies at ca. 1650 m a.s.l., but falls down to -125 m below the sea level in the northeast, in the Awash Valley (Danakil depression). The Western Ethiopian highlands receive maximum rainfall (heavy rains coming from the west Atlantic front mainly from June to September, or almost all year round in the far west of the country), while the Eastern Ethiopian highlands receive monsoon precipitations coming from the Indian Ocean in two rainy seasons due to seasonal variation of the ITCZ (Diro et al., 2011). Rainfall is unimodal in the west of the Rift valley and bimodal everywhere else (Diro et al., 2009, 2008) (**Figure 2.2**). Dry season, expressed here as the number of months with <30 mm rainfall lasts just three months in the western regions of Amhara and Benishangul-Gumuz Regional States, but is as long as 6 months in the Afar Region and Somali Regional State (Fick and Hijmans, 2017; Friis et al., 2010a). In Ethiopia, temperature decreases with increasing elevation according to a thermal gradient of about  $-0.4^{\circ}\text{C}/100\text{ m}$ . In the rift valley, mean annual temperature is about  $30^{\circ}\text{-}34^{\circ}\text{C}$  in the northernmost area of the rift valley (Danakil Depression), and about  $20^{\circ}\text{-}23^{\circ}\text{C}$  in the Great Rift Lakes Region. Mean annual temperature in Ethiopia is as low as  $6\text{-}7^{\circ}\text{C}$  above 4100 m a.s.l. where freezing is common.

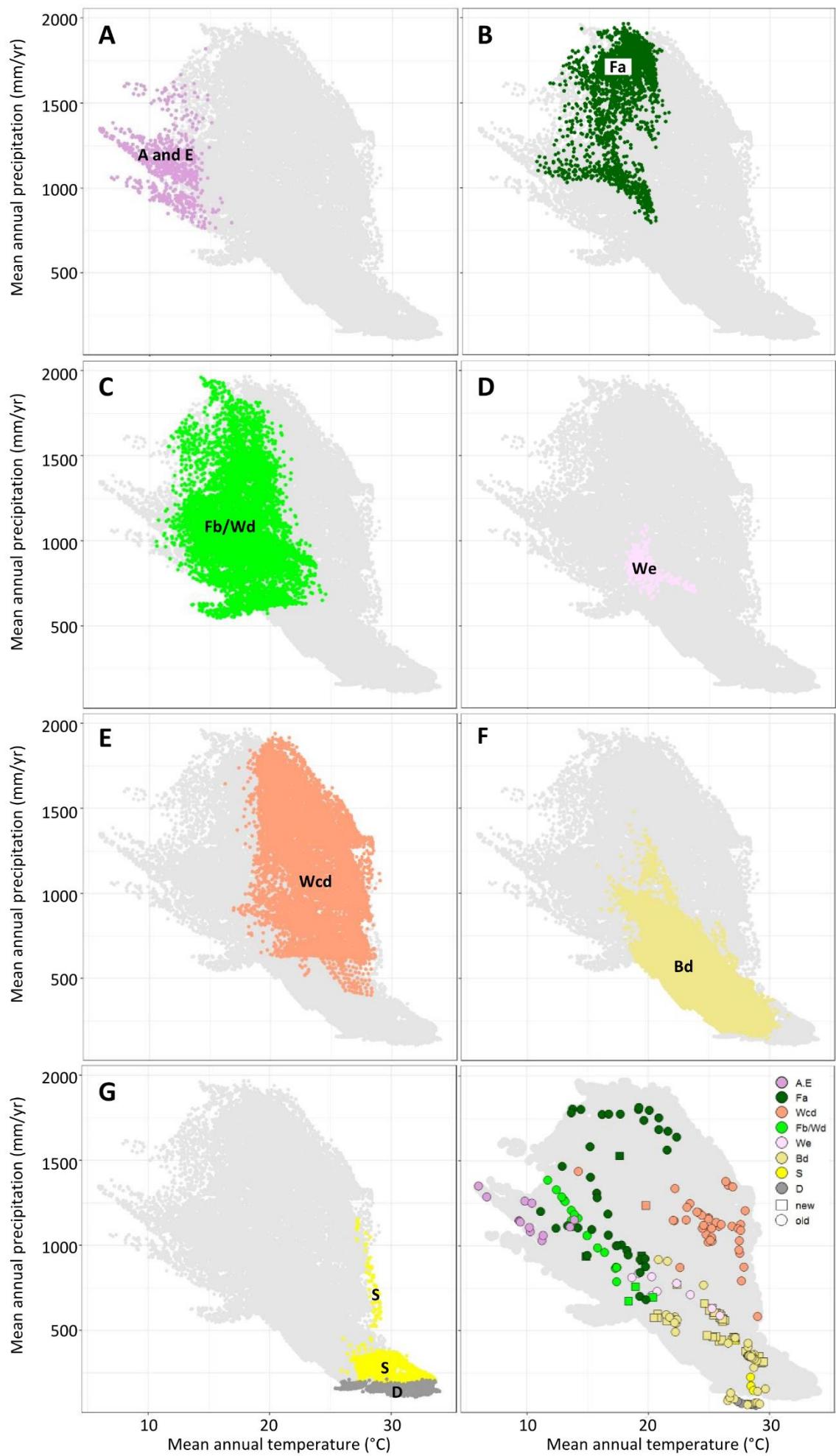


**Figure 2.1** – Distribution of pollen samples from this study. The red circles are the old samples and the white squares are the new samples.



**Figure 2.2** - Rainfall patterns for Ethiopia based on figures from Diro et al. (2008). Annual cycle for the homogenous rainfall zones of Ethiopia. The annual cycle was averaged over the 35 year period (1969–2003). The annual cycle in broken lines represent stations added to the zone after considering the cross correlation criteria. The solid lines represent stations which are retained in the zones after considering the cross correlation criteria. Different symbols indicate different stations in the zone. The pollen samples used in this study are positioned on the map with their post-hoc inference of regional vegetation and information on their edaphic context (see Table 2.1 for abbreviations).

Vegetation composition and physiognomy are largely influenced by precipitation and temperature. Friis et al. (2010) describe 12 main regional vegetation types and 12 vegetation sub-types for Ethiopia based on dominant species, vegetation physiognomy, and environmental parameters (eg., soil, climate, elevation). A georeferenced map of vegetation types and sub-types over Ethiopia and Eastern Africa (with homogenized names and abbreviations) was produced under the Vegetation and Climate change in Eastern Africa (VECEA) project (van Breugel et al., 2015), and has been used here. Vegetation distribution can be summarized as follows. Afroalpine vegetation (A) including Ericaceous zones (E) occur at high elevation (>3000 m a.s.l.) where mean annual temperature ranges between ca. 15° and 6°C (**Figure 2.3**). A complex of Afromontane forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd) occurs below the Afroalpine A and E zones and covers most of the Western and Eastern Ethiopian highlands where mean annual temperature ranges is ca 22.5-12.5°C between ca 1800-3000 m a.s.l. (but upper and lower limits can vary). Afromontane rainforest (Fa) are restricted to the south-western Highlands (Illubator and Kefa regions), where mean annual temperature ranges between ca. 12.5° to 20°C, and mean annual precipitations are of ca 1800-2000 mm/yr. At low and mid-elevation (ca. 400-1800 m a.s.l.), *Combretum-Terminalia* woodland and wooded grassland (Wcd) (with small to moderate sized trees with large deciduous leaves) occur in the west where mean annual precipitation is ca. 600-2000 mm/yr, rainfall is unimodal with a dry season of 3 to 6 months, and mean annual temperature ranges from ca. 18°C to 28°C. At similar altitude (ca. 400-1800 m a.s.l.) and comparable mean annual temperature range (ca. 18°-30°C) but with lower annual precipitations (<1000 mm/yr) and a bimodal rainfall regime, the eastern and southern regions are covered with *Acacia-Commiphora* woodlands and bushlands (Bd). In the central part of the Rift Valley below 1800 m a.s.l., where mean annual temperature is ca. 15°-24°C, rainfall is also bimodal but higher (600-1000 mm/yr), the vegetation is less xerophytic and is described as Upland Acacia wooded grassland (We). Desert (D) and semi-desert grassland and shrubland (S) occur in northernmost area of Ethiopia rift valley and eastern Ethiopia below ca 400 m a.s.l., where mean annual temperature is ca. 25°-34°C. The mean annual precipitations is < 200 mm/yr for D and generally is 200-400 mm/yr for S but can be up to 1150 mm/yr. In this most arid land, small trees, shrubs and herbs may be succulent, geophytic or annual. We note that boundaries of main vegetation types match the bioclimatic



**Figure 2.3** - Present-day distribution of the vegetation types in the climate domain of Ethiopia. A-H, Climate domain of Ethiopia (grey shading) with each of the 53381 dots representing  $4.63 \times 4.63$  km) defined by mean annual precipitation (mm/yr) (WordClim2, 1970-2000, Fick and Hijmans, 2017). A-G, Climatic domain occupied by the main vegetation types of Ethiopia obtained using the georeferenced map of vegetation types and sub-types over Eastern Africa produced under the Vegetation and Climate change in Eastern Africa (VECEA) project (van Breugel et al., 2015). A: Afroalpine vegetation ; Bd: *Acacia-Commiphora* deciduous bushland and thicket ; D: Desert ; E: Ericaceous belt ; Fa: Afromontane rain forest ; Fb/Wd: Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket ; S: Semi-desert grassland and shrubland ; Wcd: Dry *Combretum* wooded grassland (Wcd) ; We: Upland *Acacia* wooded grassland. H, Position of our 249 surface samples within the present-day climatic domain of Ethiopia and according to the post-hoc inference of regional vegetation (Table 2.1).

boundaries except in south-western Ethiopia, where the *Acacia-Commiphora* woodland and bushland (Bd) (in the east) meets with *Combretum-Terminalia* woodland and wooded grassland (Wcd) (in the west) (Friis et al. 2010) (**Figure 2.3**). Edaphic vegetation associated with rivers, freshwater lakes and salty lakes was also studied in detail and mapped by Friis et al. (2010), who distinguished up to seven types and sub-types. They noted that riverine vegetation is quite similar to vegetation in the major type above 1800 m a.s.l., but shows very different floristic composition from the surrounding vegetation below 1800 m a.s.l. (Friis et al., 2010a, p. 42). Riverine habitats below 1800 m a.s.l. may contain species found in forests under higher rainfall regimes at the same elevations. Floristic composition of riverine vegetation was estimated at 242 species. 26% of these (including 60 woody species) are exclusively or predominantly recorded in this habitat (Friis et al., 2010a, p. 127).

## 2.4. Material

### 2.4.1. Modern pollen samples and sampled climate space

This study includes 283 modern pollen assemblages obtained from 207 soil samples and 76 mud samples from Ethiopia (251), the Arabian Peninsula (17 samples from Yemen, Saudi Arabia, and Qatar) and Tanzania (15) (**Figure 2.1** and **Annexe table 2.1**). Out of these, 221 assemblages were extracted from the African Pollen Database (APD) ("African Pollen Database," 2019) (128) or the literature (93), and 62 samples (22% of the dataset)

are new. They were collected in 2009 by D. Barboni and R. Bonnefille (Ethiopian samples DB09) and in 2017 by D. Barboni, B. Bourel and J.-C. Mazur (Ethiopian samples BB17) in two areas of southern and northeastern Ethiopia not yet documented with *Acacia-Commiphora* deciduous bushland and thicket (Bd), Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd), and Afromontane rain forest (Fa) (**Figure 2.1**). Some samples labelled BB17 also document spring areas in the Awash River valley. To improve representation of spring-associated wetlands and forests, 15 samples from Tanzania were added (Barboni et al., 2019). Samples from the Arabian Peninsula (Bonnefille and Riollet, 1988; El-Moslimany, 1983; Lézine et al., 1998) were included to represent very arid conditions (annual precipitation < 125 mm/yr) that are not currently found in the modern Ethiopian Rift Valley.

This dataset samples most of the climate domain of Ethiopia, except where annual temperature is above 30°C (ie, the Danakil Depression) (**Figure 2.3**). Our new samples fill several gaps. First, they document the driest range of the *Acacia-Commiphora* deciduous bushland and thicket (Bd), which occurs under 400-600 mm/yr precipitation within the 20°-30°C mean annual temperature zone. Second, they document the hottest and driest range of occurrence of the Complex of afromontane forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd) where mean annual temperature is 18-21°C and mean annual precipitation 700-800 mm/yr (**Figure 2.3**). Despite these additions, sampling of the climate spaces could still be improved for Afroalpine vegetation and Ericaceous belt (A.E), Dry *Combretum* wooded grassland (Wcd), Complex of afromontane forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd), and Semi-desert grassland and shrubland (S) and Desert (D) (**Figure 2.3**).

#### **2.4.2. Geopositioning and assignment of environmental data**

To make local vegetation reconstructions possible, geolocation of some pollen samples required to be reworked. This was especially true for samples that were collected at a time when GPS was not available or not very accurate (Bonnefille, 2018). To do so, we used all material available: authors' notes, descriptions of vegetation at sampling sites, mission preparation documents, road maps, elevation data obtained with altimeters, and

mission reports, in addition to information provided in the original publications (Barboni et al., 2019; Bonnefille et al., 1987, 1993; Bonnefille, 1972; Bonnefille et al., 2004; Bonnefille and Buchet, 1986; Bonnefille and Riollet, 1988; El-Moslimany, 1983; Garnier et al., 2013; Lézine, 1981; Lézine et al., 1998; Mohammed, 1992; Mohammed et al., 2007; Mohammed and Bonnefille, 2002) (column 10 in **Annexe table 2.2**). For example, in the Hadar area (north of the Awash valley), samples E010 to E026 are archived under the same geographical coordinates in the APD, although author's notes and a map in the original reference show that they were collected at different locations along the Awash River, and not some 10 to 16 km away in the xerophytic bushland (Figure 5 in Bonnefille et al., 1987). For the purpose of our study, their geographical coordinates were therefore corrected (**Figure 2.4**). In total, we corrected the geographical coordinates for 68% of the "old" pollen samples (150/221 pollen samples) with an average correction of 9.7 km (**Annexe table 2.1**).



**Figure 2.4** - Comparison of the corrected geographical coordinates for the pollen samples E010 to E026 (orange square) with the original geographical coordinates (white square).

Using a geographic information system software (GIS QGis, v3.2.2), we assigned to all 283 sites (with corrected coordinates) 12 soil parameters (SoilGrids, 2000-2015,  $0.25 \times 0.25$  km, Hengl et al., 2017), 39 climatic parameters (WorldClim2, 1970-2000,  $0.93 \times 0.93$  km, Fick and Hijmans, 2017), five aridity parameters (Global High-Resolution Soil-Water Balance, 1950-2000,  $0.93 \times 0.93$  km, Trabucco and Zomer, 2019), Net Primary Productivity (NPP) (MODIS MOD17A3, 2000-2014,  $0.93 \times 0.93$  km, Zhao et al., 2005), and percent of tree cover (Global Maps, year 2008,  $0.46 \times 0.46$  km, Kobayashi et al., 2017) (**Annexe table 2.3**). Satellite percent tree cover values for the 2000-2014 period were sometimes very different from the authors' notes at the time of sampling (some 40 years

ago in some cases). Coherence between modern satellite tree cover values and authors' notes was therefore checked for every site, and when deforestation was observable on Google Earth images (columns 12 and 14 in **Annexe table 2.2**), then tree cover values of neighbouring, unaffected or less affected areas within 1-kilometer radius of the sampling point were considered. Similarly, we corrected tree cover values when a slight discrepancy (<250 m) was observed between Google Earth images and the tree cover grid. In total, tree cover values were corrected for 116 sites (over 283). GIS was also used to attribute to each surface sample (with corrected geographical coordinates) a potential vegetation type according to the VECEA project (van Breugel et al., 2015) for African samples and according to phytogeographical map of Frey and Kürschner (1989) for Arabian samples (columns 9 in **Annexe table 2.2**). For ca. 80 samples of our dataset the assigned VECEA vegetation type did not match the description of vegetation given by the collectors. We gave priority to the collector's notes and vegetation observed with Google Earth to correct the assigned potential vegetation accordingly (**Annexe figure 2.1**).

## 2.5. Analytical approach

### 2.5.1. Original pollen data and transformations

Pollen taxa names were homogenized according to Vincens et al. (2007) (**Annexe table 2.4**). Pollen types are difficult to differentiate in the modern samples and rarely differentiated in the fossil register were grouped together (eg., *Nuxia/Dobera*). Taxonomic precision was limited at the subfamily level for Asteraceae. Spores were excluded from our analyses because it is not clear if they were consistently identified and counted by all analysts. *Prosopis*-type and *Prosopis*-type *africana*, which exhibit little morphological differences were also removed from the dataset (after the calculation of percentages), because *Prosopis*-type is produced by the recently introduced and very invasive species *Prosopis chilensis*.

Data transformations and analyses were carried out using R (R Development Core Team, 2018). Pollen percentages were calculated based on total pollen sum, spores and undetermined pollen excluded (raw data in **Annexe table 2.5**). The double entry matrix of  $n$  pollen taxa x  $m$  sites with percentages was then transformed into a column vector  $\vec{V}$ .

Classified in ascending order, values of  $\vec{V}$  show a break at 0.2% in the evolution of pollen percentages. We interpret this break as the signal of allochthonous taxa, which occurrence in pollen assemblages is erratic. Before further processing of the data, all  $\vec{V}$  entries with values <0.2% were therefore excluded from the column vector to reduce the noise associated with the presence of allochthonous pollen (**Annexe figure 2.2**). In other words, we chosen a threshold pollen percentage of 0.2% for this study.

Pollen percentages were then converted into an abundance scale to qualify differences in pollen percentages between samples from similar vegetation types. To do so, after removal of pollen percentage < 0.2% in each pollen sample, we used the Vegan function `cascadeKM()` (Vegan v2.5-2: Oksanen et al., 2010) that applies the *k*-means partition comparison method and the simple structure index (ssi) criterion to identify the minimum number of classes of abundances that keep a maximum of information (Dolnicar et al., 1999). The larger the ssi, the more significant is the partitioning (Borcard et al., 2018). Based on this index, 2 to 6 classes were optimal (**Annexe figure 2.3**). We retained six classes. To define the classes we used the Fisher-Jenks method of the Cartography function `getBreaks()` (Cartography v2.2.0: Giraud and Lambert, 2017): abundance class 1  $\in ]0.2\% ; 3\%]$ , abundance class 2  $\in ]3\% ; 11\%]$ , abundance class 3  $\in ]11\% ; 23\%]$ , abundance class 4  $\in ]23\% ; 38\%]$ , abundance class 5  $\in ]38\% ; 60\%]$  and abundance class 6  $\in ]60\% ; 100\%]$ . A 7<sup>th</sup> class was added for the threshold pollen percentage of 0.2%: abundance classes 0  $\in ]0\% ; 0.2\%]$ . It is the application of these 7 abundance classes on the pollen percentages (without removal of pollen percentages < 0.2%) that was used to convert the pollen data into an abundance scale.

### **2.5.2. Data mining through clustering and vegetation types identifiable using pollen assemblages**

In order, to find out what useful information could be extracted from our modern pollen dataset we used a method based on a *k*-means partitioning constrained by a hierarchical clustering. Pollen taxa occurring in less than 2% of the sample dataset (ie., in less than 6 samples) were excluded from this analysis to reduce the influence of rare types. In details, we used the normalized matrix of pollen classes to compute a chord distance matrix (Vegan function `vegdist()`, Euclidean method) for generating a hierarchical dendrogram (Stats function `hclust()`, `ward.D2` method) (Stats v3.5.1: R

Development Core Team, 2018). Function `cascadeKM()`, which starts from the centroids of the hierarchical dendrogram groups was then used to test up to 35 clusters. Based on the analysis of the ssi criterion, 21 groups appeared as the optimal number (hereafter named km-groups) (**Annexe figure 2.4**). This function of *k*-means partitioning was used to optimize the result of the hierarchical grouping, by allowing samples to be transferred from one group to the other. *K*-means partitioning allows an object that has been included in one group to being transferred to another that appeared later in the agglomeration process, if the latter is more appropriate (Borcard et al., 2018) (see **Annexe figure 2.5**). We further tested the coherence of the km-groups by analyzing their silhouettes represented by the ssi of each sample in each group (Stats function `silhouette()`) (**Annexe figure 2.6** and **Annexe table 2.2**). Negative ssi values for a group or a sample are interpreted as misclassification (Borcard et al., 2018), so we removed from further analyses all samples with negative ssi values, as well as the 10 samples from km-group #11 which average silhouette width was negative. To improve the classification, we also removed sample E234 which, despite positive ssi values, is the only representative for the afromontane bamboo vegetation sub-type. At last, sample E212 was reclassified from km-group #4 (riparian) to #18 (terrestrial *Combretum* woodland) because of its very high percentage of Poaceae (>80%) which gives it an aquatic rather than terrestrial signal (**Annexe table 2.2**). In total, 34 samples were removed (12% of the modern dataset) and one sample was reclassified. The remaining 249 samples are clustered into 20 valid km-groups (#1 to #21, #11 excluded), each of these 20 km-groups consists of a set of samples having a homogeneous pollen signal and significantly different from the pollen signal of the other groups. The distribution of individual samples in the km-groups is given in **Annexe table 2.2**, but a summary by major vegetation types is given in **Table 2.1**.

The vegetation types represented by the km-groups were inferred post-hoc from 3 observations: (1) the distribution of the samples in the km-groups (**Table 2.1**), (2) the potential natural (regional) vegetation (Frey and Kürschner, 1989; van Breugel et al., 2015), and (3) the (local) vegetation described at the sampling sites (see references in section 3.2, and column 15 in **Annexe table 2.2**). We also used other publications not directly related to our samples to fill in the edaphic data (eg., spring) (Ayenew et al., 2008; Cuthbert et al., 2017). The km-groups were labelled using abbreviations, including those proposed by Friis et al. (2010) and used in the VECEA project (van Breugel et al., 2015) that relate to regional vegetation types. All 20 km-groups can be associated with only one

type of vegetation and are mainly constituted of soil samples, except km-groups #04 and #15, which are exclusively constituted of mud samples collected in river beds (**Table 2.1**). Km-group #04 includes mud samples from rivers going through riparian vegetation in upland Acacia wooded grassland (We), *Acacia-Commiphora* deciduous bushland and thicket (Bd), and Semi-desert grassland and shrubland (S), while km-group #15 includes samples going through riparian vegetation in afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd) and Dry *Combretum* wooded grassland (Wcd) (**Table 2.1**).

1. Regional vegetation types identified by botanists <sup>0,2</sup>	2. Sample distribution in the different k-means groups															3. Depo.	4. Post-hoc inference of vegetation types (◦) and sub-types (*)				
	21	20	19	05	08	03	06	07	17	12	01	02	18	10	13	16	14	09	04	15	
Desert (D)																				soil	• Desert (D) <sup>0,2</sup>
Semi-desert grassland and shrubland (S)																				soil	• Xerophytic dwarf shrubland <sup>3</sup> (S-xds)
																				soil	• Xerophytic shrubland <sup>3,6</sup> (S-xs)
																				mud	• Riparian very open woodland/shrubland along meandering perennial or seasonal river in S <sup>2,4,8</sup> (S-river)
Acacia -Commiphora deciduous bushland and thicket (Bd)																				soil	• Riparian dense <i>Tamarix</i> forest <sup>2,6</sup> (Bd-rtf)
																				soil	• Spring-fed woodland with <i>Acacia</i> , <i>Hyphaene</i> and <i>Tamarix</i> <sup>2,4,5</sup> directly on springhead (Bd-spring 1)
																				soil	• Spring-fed woodland with <i>Acacia</i> , <i>Hyphaene</i> or <i>Phoenix</i> <sup>4,5</sup> not directly on springhead (Bd-spring 2)
																				soil	• Steppe-like bushland with <i>Boscia</i> <sup>2,4</sup> (Bd-wb 1)
																				soil	• Open woodland/bushland with <i>Amaranthaceae</i> and <i>Ocimum</i> <sup>2,4</sup> (Bd-wb 2)
																				soil	• <i>Acacia</i> woodland and bushland <sup>2</sup> above 550m <sup>4,6</sup> (Bd-wb 3)
																				soil	• Moist edaphic <i>Acacia</i> woodland with abundant grasses and Cyperaceae <sup>2,3,4,5</sup> (Bd-wgc)
																				mud	• Riparian open <i>Acacia-Tamarix</i> forest surrounded by swamp along meandering river in Bd and upper limit of S <sup>2,4,8</sup> (Bd-river)
Upland Acacia wooded grassland (We)																				mud	• Riparian <i>Acacia</i> forest along straight river in We and upper limit of Bd <sup>2,4,8</sup> (We-river)
Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd)																				soil	• Fb/Wd below 2500m without incursion of Ericaceous belt <sup>2</sup> (Fb/Wd-low/middle)
																				soil	• Fb/Wd above 2600m with incursion of Ericaceous belt <sup>2</sup> (Fb/Wd-upper)
																				mud	• Riparian afromontane undifferentiated forest along straight river in Fb/Wd <sup>2,4,8</sup> (Fb/Wd-river)
Dry <i>Combretum</i> wooded grassland (Wcd)																				soil	• Dry <i>Combretum</i> wooded grassland (Wcd) <sup>0,2</sup>
																			soil	• Riparian open <i>Acacia</i> forest along straight river in Wcd <sup>2,8</sup> (Wcd-river 1)	
																			mud	• Riparian close <i>Acacia-Cordia</i> forest along meandering river in lower limit of Wcd in contact with Bd <sup>2,8,9</sup> (Wcd-river 2)	
Afromontane rain forest (Fa)																				soil	• Dry Fa <sup>2,7</sup> with PANN < 1100mm/yr (Fa-sc 1)
																			soil	• Moist Fa <sup>2,7</sup> with PANN > 1100mm/yr (Fa-sc 2)	
																			soil	• Secondary moist evergreen afromontane forest <sup>2</sup> (Fa-w)	
																			soil	• Upper transitional montane forest below E <sup>7</sup> (Fa-upper)	
Afroalpine vegetation and Ericaceous belt (A.E)																			soil	• <i>Erica</i> bushland <sup>2</sup> or grove within <i>Hilechrysum</i> heath <sup>2,7</sup> (A.E-e)	

**Table 2.1** – Regional vegetation, depositional environments, vegetation types and sub-types associated with the pollen samples present in each k-mean groups. 0 : van Breugel et al. (2015). Vegetation types and sub-types are based on 1: El-Moslimany (1983), 2: Friis et al. (2010), 3: Frey and Kürschner (1989), 4: *in situ* observations, 5: Barboni et al. (2019), 6: Bonnefille et al. (1987), 7: Mohammed et al. (2007), 8: Buchet (1982), and 9: Carr (1998).

To take into account the vegetation diversity observed in riparian vegetation from different regions (Friis et al. 2010) and since km-groups #4 and #15 had large sample sizes (67 in total), we artificially created six additional *river* groups: S-river, Bd-river, We-river, Fb/Wd-river, Wcd-river 1, and Wcd-river 2, and ran a new clustering analysis. These groups were created by considering initial assignment of samples to km-group #4 or #15, sample location (elevation and vegetation zone), and river type (meandering or non-meandering) (see **Table 2.1** and **Annexe table 2.2** for details). These six groups of rivers are each associated with a type of local riparian vegetation in dynamic depositional environments (**Table 2.1**)

### 2.5.3. Vegetation score for the km-groups (V score)

Pollen taxa affinities for each km-group and, by extension each vegetation, were estimated with the Value Test (VT) (Lebart, 2000). VT compares for one variable (pollen taxon) the values of descriptive statistic indicators (means, variances, and group size) computed on all groups (the full dataset) and on a subsample (a km-group) related to this group. The VT of a pollen taxon for a given km-group is calculated as follows:

$$VT = \frac{\mu_g - \mu}{\sqrt{\frac{n-n_g}{n-1} \times \frac{\sigma^2}{n_g}}}$$

where  $\mu_g$  is the mean abundance of a given pollen taxon within a given km-group,  $\mu$  the mean abundance of the pollen taxon within the whole dataset,  $n_g$  the size of the group (number of samples in a given km-group),  $n$  the size of the whole dataset (total number of samples), and  $\sigma^2$  the variance associated with mean abundance of a given pollen type in the whole dataset (Lebart et al., 1995). The VT of a pollen taxon for a km-group is a simple measure of similarity between the pollen taxon and the group (km-group, and by extension a vegetation type). Similarity is indicated by positive VT, and dissimilarity by negative VT (columns 2-21 in **Annexe table 2.6**).

Using VT values of the pollen taxa for the km-groups, we then developed the calculation of a score to measure the affinity of the pollen samples for each of the 20 km-groups. The affinity of a pollen assemblage to a given km-group is calculated as follows, and then transformed into a percentage:

$$A_{\alpha,\beta} = \frac{\sum_{i=0}^9 \left( \sum P_{VT\% \in [\frac{10i \times VT_{max}}{100}, \frac{10(i+1) \times VT_{max}}{100}]} \times i \right)}{\sum_{i=0}^9 \left( N_{VT\% \in [\frac{10i \times VT_{max}}{100}, \frac{10(i+1) \times VT_{max}}{100}]} \times i \right)} - \frac{\sum_{i=0}^9 \left( \sum P_{VT\% \in [\frac{10(i+1) \times VT_{min}}{100}, \frac{10i \times VT_{min}}{100}]} \times i \right)}{\sum_{i=0}^9 \left( N_{VT\% \in [\frac{10(i+1) \times VT_{min}}{100}, \frac{10i \times VT_{min}}{100}]} \times i \right)}$$

where  $A_{\alpha\beta}$  is affinity of sample  $\beta$  for the km-group  $\alpha$ ,  $P_{VT\% \in ]x; y]}$  is the abundance scale of a pollen type in sample  $\beta$  with VT percentage between  $]x; y]$  for the km-group  $\alpha$ ,  $N_{VT\% \in ]x; y]}$  is the number of pollen types in the sample  $\beta$  with VT percentage between  $]x; y]$  for the km-group  $\alpha$ . The VT percentage of a pollen type for the km-group  $\alpha$  is  $[VT_{pollen\ type\ X} \times 100]/VT_{max}$ ,  $VT_{max}$  is the most positive VT of pollen type for the km-group  $\alpha$ , and  $VT_{min}$  is more negative VT of pollen type for the km-group  $\alpha$ . In this formula, the first member of the formula (left) is the weighted mean of the abundance scale of all pollen types in sample  $\beta$  with a positive VT for the km-group  $\alpha$ . The weight of each abundance scale is based on the VT percentage of its pollen type such as:  $VT\% \in ]10n ; 10(i+1)]$  for weight= $i+1$ . The second member of the formula (right) does the same thing but for negative VT.  $A_{\alpha\beta}$  therefore represents the weighted mean of pollen of sample  $\beta$  with a similarity for the km-group  $\alpha$  minus the weighted mean of pollen of sample  $\beta$  with a dissimilarity for the km-group  $\alpha$ . With a value of  $A_{\alpha\beta} < 0$ , it is the dissimilarity signal which is predominant on the similarity signal. Put in another way, the marker pollen signal of the km-group  $\alpha$  is hidden by the unmarker pollen signal of the km-group  $\alpha$ .  $A_{\alpha\beta} < 0$  is therefore a score mainly based on the absence of marker pollen.

In order to have a Vegetation score (V score) related more to the presence of pollen markers than to their absence, we first set to 0 all negative  $A_{\alpha\beta}$ , and then transformed the A score such as the final V score is expressed as follows:

$$V_{\alpha,\beta} = \frac{A_{\alpha,\beta} \times 100}{A_{total,\beta}}$$

where  $V_{\alpha,\beta}$  is the final score of sample  $\beta$  for the km-group  $\alpha$  (expressed between 0 and 100),  $A_{\alpha,\beta}$  is affinity of pollen sample  $\beta$  for the km-group  $\alpha$ ,  $A_{total,\beta}$  is the sum of all A scores of sample  $\beta$ . The V score of a sample for a km-group is the percentage of affinity score (A score) of this sample for this km-group (and by extension the vegetation type that was allotted to him) on the sum of all A scores of this sample.

#### **2.5.4. Vegetation score in dynamic depositional environment (V-river score)**

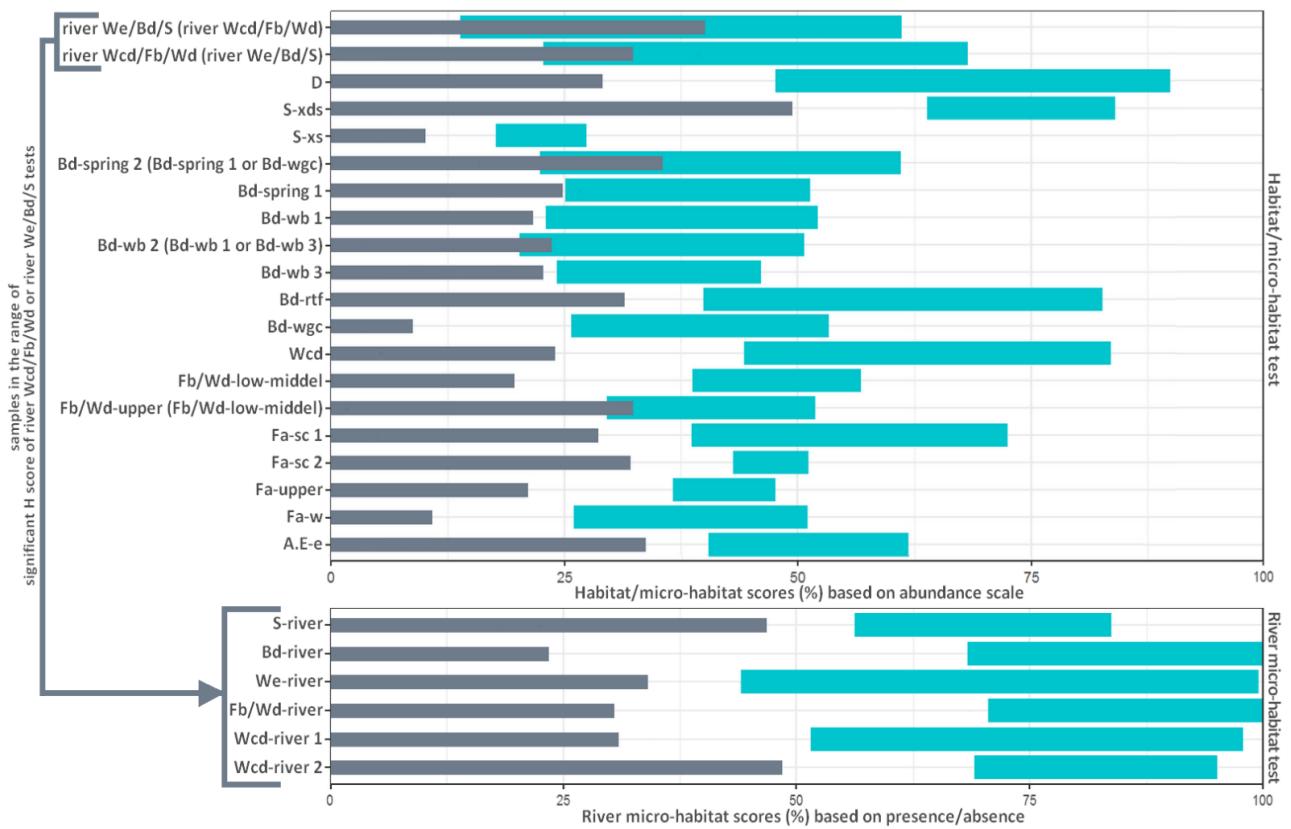
For the six *river* groups composed of the 67 mud samples initially clustered into km-groups #4 and #15 (part 5.2), the V scores were calculated. This variation of V score for these six types of local riparian vegetation in dynamic depositional environments is called V-river score. After observing that the mud samples differed drastically from the soil samples in terms of pollen diversity rather than pollen abundance, use pollen data in presence/absence seemed more appropriate than in abundance scale. We decided to calculate the V-river in two ways for comparison: presence/absence and abundance scale.

#### **2.5.5. Significant and non-significant score ranges for V and V-river scores**

For a selected km-group, we calculate the V score for this km-group for all pollen samples. The V score values of the samples classified in this km-group (**Table 2.1**) define the range of significant V score values for this km-group with a confidence interval of 95% (cyan bar in **Figure 2.5**). The V score values of the samples that are not classified in this km group define the range of non-significant V score values for this group with a 95% confidence interval (grey bar in **Figure 2.5**). The same procedure was applied for the significance or non-significance of a V-river score value for the *river* groups (**Figure 2.5** and **Annexe figure 2.7**).

#### **2.5.1. Environmental parameters for post-hoc inferred vegetation**

To see which environmental, climatic, and/or edaphic variables would best relate to the 24 final groups (the 20 km-groups except #4 and #15 which are replaced by the six *river* groups.), and by extension the associated 24 vegetation types and sub-types (**Table 2.1**), we ran a Random Forest model of 10000 iterations with the function Random Forest () of the Random Forest v4.6 R package. We used 10 variables measuring temperature, 10 for precipitation, 5 for aridity, 11 for soil, tree cover percentage, and net primary production (NPP) (columns 10-28 and 40-59, **Annexe table 2.3**). We then applied the function varImpPlot () of the same R package with our model to see which variables best explain the distribution of vegetation with the function.



**Figure 2.5** - Range of score values significant (cyan) and non-significant (grey) for each vegetation type and sub-type with a confidence interval of 95%. In brackets, vegetation types or sub-types with overlapping significant scores. For abbreviations, see Table 2.1.

## 2.5.2. Rainfall patterns for post-hoc inferred vegetation

Using the mean precipitation of each month (columns 29 to 40 of **Annexe table 2.3**), we calculated the monthly mean precipitation of samples in the same vegetation, and determined the seasonal rainfall pattern corresponding to the post-hoc inferred vegetation. Before calculating the means, we checked whether several rainfall patterns were present in samples of a same post-hoc inferred vegetation and we have subdivided the samples of these post-hoc inferred vegetation in function of this. For our "Hyper Arid" sites, we decided to consider a dry month when mean precipitation was <30mm/month, following Trabucco and Zomer (2019).

## 2.6. Results

Our approach allowed us to organize 88% of our dataset (249 samples) in 20 robust km-groups (**Table 2.1**). The remaining 12% of the initial dataset (35 samples) could not be robustly classified. Within these 12%, nine samples are from anthropic sites (eg., DB09-04, Karasoditi coffee forest), three samples are from very poorly sampled vegetation (eg., BB17-41, pioneer vegetation growing in small patches on 200 years-old lava flow), 10 samples are from ecotones (eg., E019, limit between riparian forest and sub-desert steppe). For 12 other samples, we could not find any reason why they couldn't be assigned to a km-group. These samples were therefore excluded from further analyses.

### 2.6.1. Comparison of km-groups with the real distribution of local and regional vegetation at the sampling sites

All 20 km-groups except km-groups #04 and #15, show that the pollen assemblages allow to distinguish several vegetation types or sub-type, which botanists also identified based on their floristic composition, and each km-group is associated with only one vegetation (**Table 2.1**). For example, km-group #16 represents only Secondary moist evergreen afromontane forest, a sub-type of Afromontane rain forest (Fa). In contrast, km-groups #04 and #15, which are almost exclusively constituted of mud samples represent riparian vegetation. These two km-groups, however, are each associated with several riparian vegetation sub-types from different regional vegetation. Km-group #04 is principally composed of mud samples from the Awash river crossing the vegetation areas of Upland *Acacia* wooded grassland (We), *Acacia-Commiphora* deciduous bushland and thicket (Bd), and Semi-desert grassland and shrubland (S). Km-group #15 is composed of mud samples from the Omo river crossing the Dry *Combretum* wooded grassland (Wcd), the Wcd in contact with *Acacia-Commiphora* deciduous bushland and thicket (Bd), and Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket (Fb/Wd). This confirms that km-groups allow (1) to discriminate the vegetation types and sub-types of dynamic depositional environments (km-groups #04 and #15) from others (other km-groups), and (2) to identify local and regional vegetation type for km-groups of non-dynamic depositional environments.

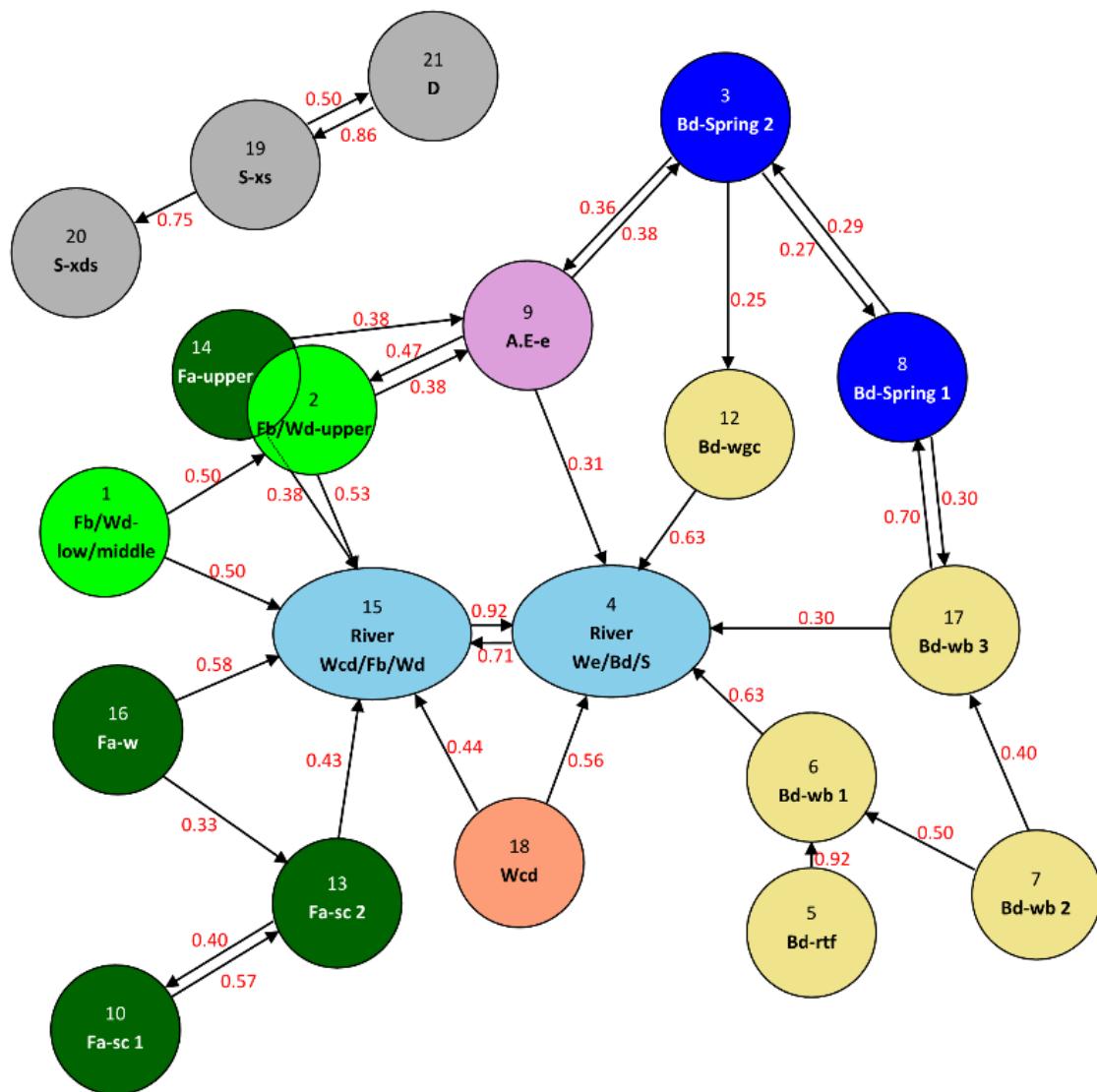
The km-group analysis assigns the most probable km-group to a sample, as well as a second, most probable group (columns 4 and 5 in **Annexe table 2.2**). This information can be used to estimate how well km-groups are related to each other. The resulting network is visible in **Figure 2.6**. The connections respect for the large majority the real phytogeographic connections (Friis et al., 2010 ; Frey and Kürschner, 1989). This strengthens the robustness of our vegetation allocations. The arid and semi-arid vegetation with km-groups #21, #20, and #19 respectively associated to Desert (D), Xerophytic dwarf shrubland (S-xds), and Xerophytic shrubland (S-xs) are isolated from all other vegetation. The km-groups are grouped according to their regional vegetation. The *river* groups (km-groups #04 and #15) are connected to most of km-groups. They have a high affinity with a large number of non-riparian groups, in accordance with botanical observations (Friis et al., 2010). The *river* groups play the role of connectors between geographically separated vegetation (eg., River-We/Bd/S connects groups Bd, Wcd, and Afromontane (A.E)). The *river* groups are characterized by very varied pollen assemblages which are representative of the regional vegetation crossed by the rivers.

### **2.6.2. Effectiveness of V and V-river scores in vegetation identification**

For most km-groups (15/20), significant and non-significant  $V_{\alpha,\beta}$  scores do not overlap. A slight overlap (<20%) occurs for  $\alpha$ =km-group #07 (Bd-wb 2),  $\alpha$ =km-group #03 (Bd-spring 2),  $\alpha$ =km-group #02 (Fb/Wd-upper), and  $\alpha$ =km-group #04 (We/Bd/S-river), but a large overlap (50%) occurs for  $\alpha$ =km-group #15 (Wcd/Fb/Wd-river) (**Figure 2.5**). For  $\alpha$ =Wcd/Fb/Wd-river the range of significant scores overlaps only with the range of We/Bd/S-river scores and *vice versa*. The use of the  $V_{\alpha,\beta}$  score is therefore appropriate to distinguish the inferred vegetation from the 20 km groups except for the km-group #04 (We/Bd/S-river) and the km-group #15 (Wcd/Fb/Wd-river). However, despite these overlaps, all samples with a V score value in the significant range for  $\alpha$ =km-groups #04 (We/Bd/S-river) or/and  $\alpha$ =km-group #15 (Wcd/Fb/Wd-river) can only be from these same groups; by extension they are necessarily from the river groups. V score is also relevant for isolating the *river* groups.

We obtained better discrimination of the 6 river groups by calculating  $V_{\text{river},\beta}$  scores (score of sample  $\beta$  identified as a sample of *river* groups by V score for the *river* groups  $\alpha$ ) with pollen counts in presence/absence. Indeed, using presence/absence

instead of pollen counts in abundance scale for calculating V-river <sub>$\alpha,\beta$</sub>  scores lead to increases discrimination of the groups as significant and non-significant scores do not overlap anymore (**Figure 2.5** and **Annexe figure 2.7**). V-river <sub>$\alpha,\beta$</sub>  scores calculated with pollen counts in abundance scale, leads to overlapping of significant and non-significant scores by 10% to 50% (**Annexe figure 2.7**). V-river <sub>$\alpha,\beta$</sub>  scores with pollen counts in presence/absence are more effective in discriminating vegetation types if pollen assemblages are from mud samples, that is from samples corresponding to dynamic depositional environments.



**Figure 2.6** - Neighborly relationships between the  $k$ -mean groups. Arrows and red numbers indicate the percentage of neighboring samples (eg., 27% of the Bd-spring 2 samples are closest neighbors to the Bd-spring 1 samples and 29% of the Bd-spring 1 samples are closest neighbors to the Bd-spring 2 samples). The length of the arrows is inversely proportional to the proximity of groups. Values below 25% have not been indicated. For abbreviations, see Table 2.1.

### **2.6.3. Relation between inferred vegetation types and sub-types and environmental parameters**

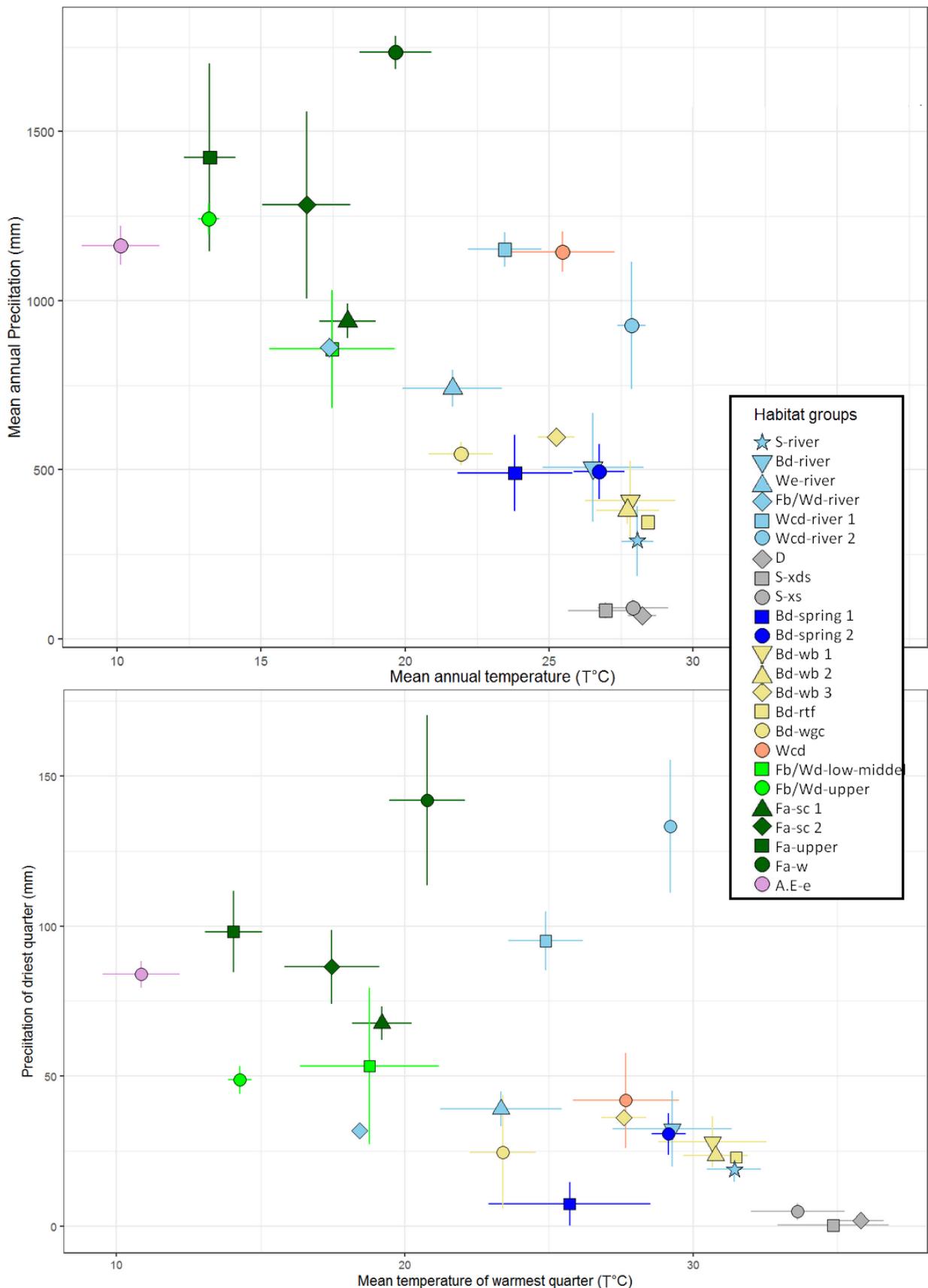
The Random Forest model was able to recreate the 24 final vegetation type and sub-types (the 20 km-groups except #4 and #15 which are replaced by the six *river* groups) (**Table 2.1**) with a success rate of 80%. This indicates that 80% of our dataset (199 pollen samples) has a pollen signal that provides information about one or several environmental parameters. The correlation between an environmental parameter and the 24 final vegetation types and sub-types is proportional to the mean decrease accuracy in Random Forest model (**Annexe figure 2.8**). Based on this, 5 parameters emerge as most correlated with vegetation types as inferred from their pollen signature. From most to least correlated, these are the tree cover percentage, mean temperature of warmest quarter, mean annual precipitation, mean annual temperature, and mean precipitation of driest quarter. All other climatic parameters including the 5 aridity parameters are only moderately correlated, the most informative aridity parameter is the annual amplitude of alpha (**Annexe figure 2.8**). The 12 soil parameters and NPP are poorly correlated.

The 24 final vegetation type and sub-types as inferred from their pollen signal clearly occupy distinct climate spaces (**Figure 2.7**) and correspond to distinct physiognomic types of vegetation with different tree cover (**Figure 2.8**). This is also consistent with the geographical distribution of the samples (**Figure 2.9**). Based on our confidence intervals of 95%, our approach allows reconstructing these climatic parameters with very good resolution of 2.5 °C for mean temperature of warmest quarter, 18 mm/yr for mean precipitation of driest quarter, 2.2 °C for mean annual temperature, and 171 mm/yr for mean annual precipitation. Looking at annual reports (mean annual temperature by mean annual precipitation) and extreme trimesters (mean temperature of warmest quarter by mean precipitation of driest quarter) helps to understand what differentiates our pollen inferred vegetation in terms of climate. This is the case for Fb/Wb-upper which is in the same range of temperature and precipitation as Fa-upper but differs from Fa-upper by much lower precipitation of driest quarter.

The climatic spaces defined by the four main climatic parameters (mean temperature of warmest quarter, mean annual precipitation, mean annual temperature, and mean precipitation of driest quarter) shows that the vegetation sub-types within a

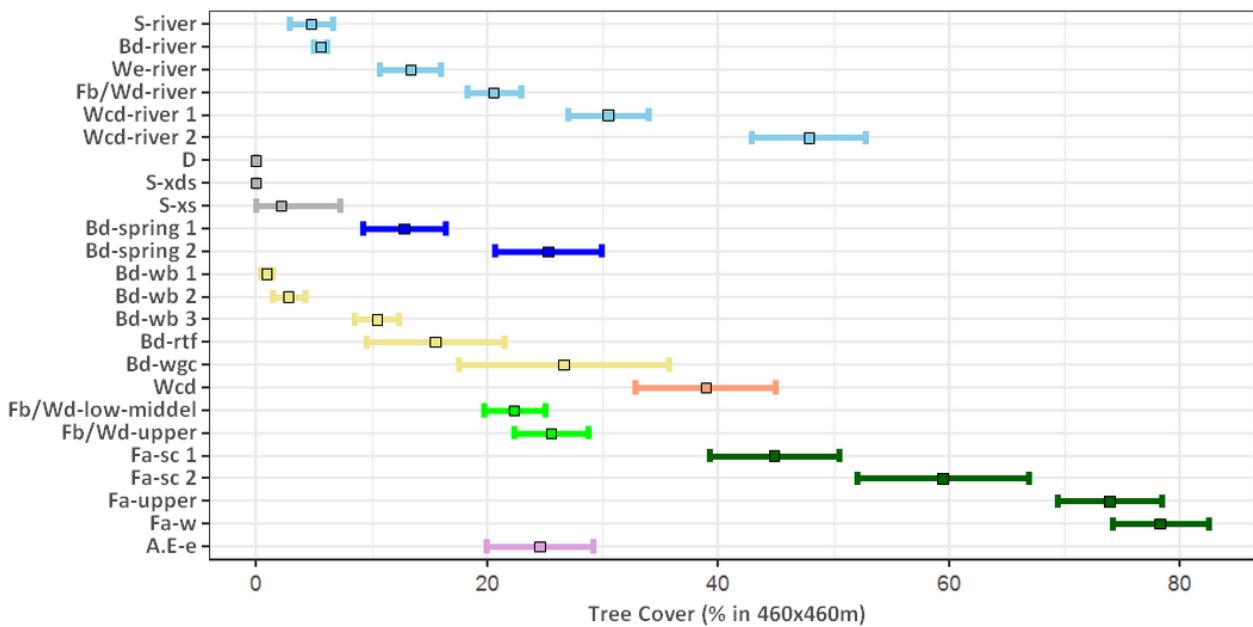
given vegetation type are usually complementary and barely overlap (eg. Fa-sc 1, Fa-sc 2, Fa-upper, and Fa-w : **Figure 2.7**). Some overlap is observed, however, for river vegetation sub-types (eg. Bd-river), which unsurprisingly occupy similar climatic spaces as regional vegetation types (Bd-wb 1, Bd-wd 2, and Bd-wd 3 : **Figure 2.7**). Bd-wd 1 and Bd-wb 2 have very close climatic ranges that are not explained by edaphic differences. Similarly, D, S-xds, and S-xs, which have distinct pollen signals, seem to occur within the same climatic spaces as defined by the four main climatic parameters (**Figure 2.7**). We have manually checked all other environmental and climatic parameters for these sub-types, and found that D differs from S-xds and S-xs are differentiated by their precipitation seasonality (coefficient of variation of mean annual precipitation) (**Annexe figure 2.9**). But no explanation was found for Bd-wd 2 and Bd-wb 1.

Altogether, our modern pollen dataset and our newly developed approach allow vegetation inferences with very distinct ecological requirements. For example, 10 different vegetation types and sub-types can be distinguished within the 25-30°C range for mean annual temperature over a precipitation gradient ranging from as low as 125 mm/yr to almost 1250 mm/yr (**Figure 2.7**). This will allow, in the fossil record, to follow or highlight climatic variations related to changes in precipitation but without significant changes in temperature. The reverse is also feasible as shown by the 10 types and sub-types of vegetation within the 500-1000 mm/yr range for mean annual precipitation over a temperature gradient ranging from 15°C to 30°C (**Figure 2.7**). Using an extended modern climate dataset with many variables has made it possible to identify subtle differences in the climate that may be responsible for subtle changes in the floristic composition of the vegetation and, therefore, in the pollen assemblages. The 24 final vegetation types and sub-types inferred here from the pollens allow a fine monitoring of the evolution of the climatic conditions whether it is a simultaneous change of temperature and precipitation or only one of these two parameters.

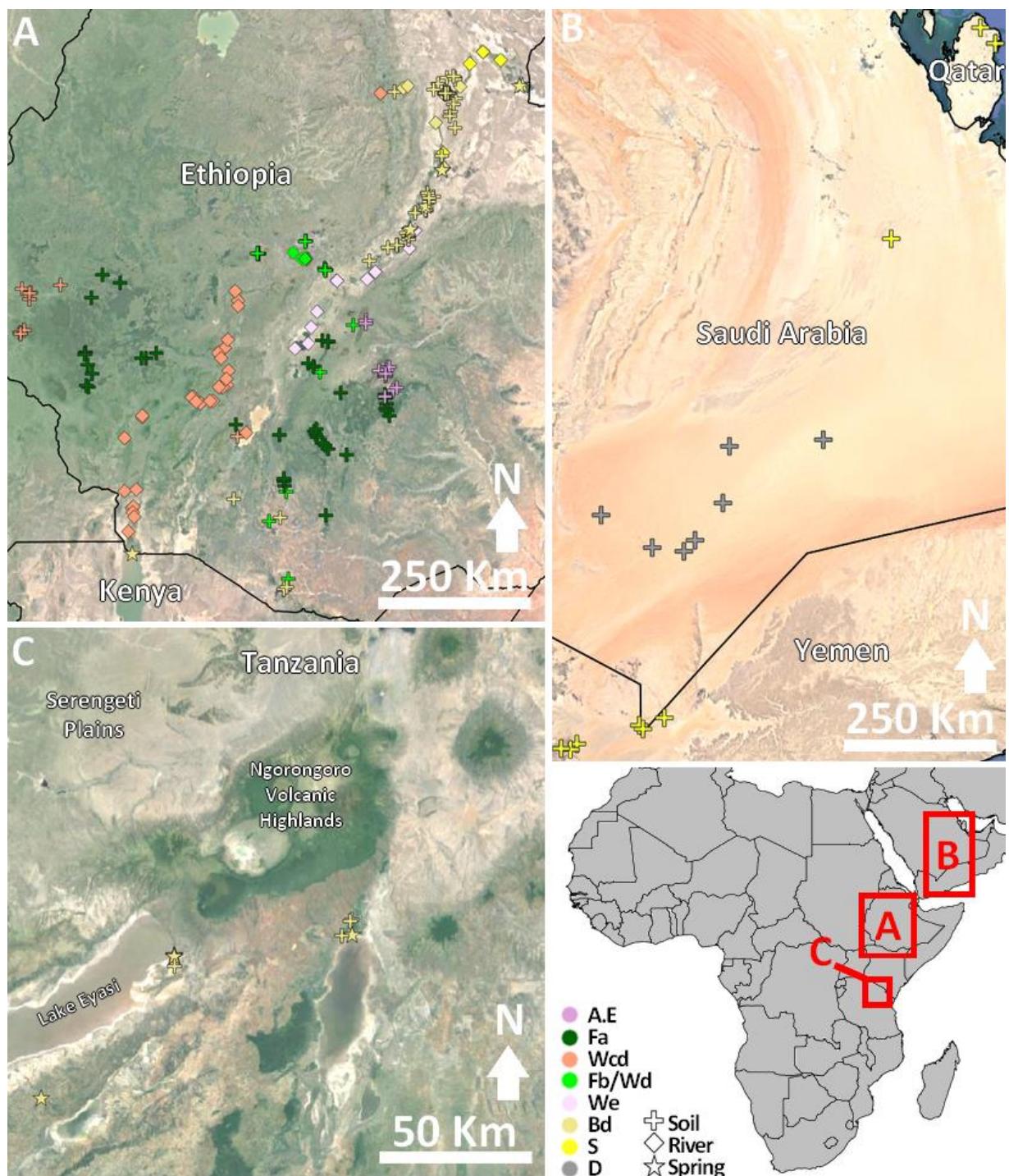


**Figure 2.7** - Ranges of mean annual temperature by mean annual precipitation and mean temperature of warmest quarter by mean precipitation of driest quarter for the post-hoc inferred vegetation represented by the samples (WordClim2, 1970-2000, 0.93x0.93 km, Fick and Hijmans, 2017) with a confidence interval of 95%. For abbreviations, see Table 1.

Regarding the tree cover, our 24 final vegetation type and sub-types make it possible to infer a wide range of tree cover from 0% to 82% (**Figure 2.8**). The tree cover percentage range identified for each vegetation has a good interval resolution with an average difference of 7% of tree cover between the maximum and minimum estimated values. Our estimated intervals of tree cover percentage are different and complementary at the scale of vegetation sub-types (cf. Fa-sc 1 and Fa-sc 2: **Figure 2.8**). Tree cover for riparian vegetation types does not exceed 52% because riparian vegetation types as identified here by their pollen signal are only distinct from the surrounding vegetation below 1800 m a.s.l.. Above this altitudinal threshold, rainfall is high enough in Ethiopia to trigger the development of woodlands and forests with >50% tree cover (Friis et al., 2010a, p. 36).



**Figure 2.8** - Range of Percent tree cover for the post-hoc inferred vegetation represented by the samples (Global Maps, year 2008, 0.46×0.46 km, Kobayashi et al., 2017) with a confidence interval of 95%. For abbreviations, see Table 1.

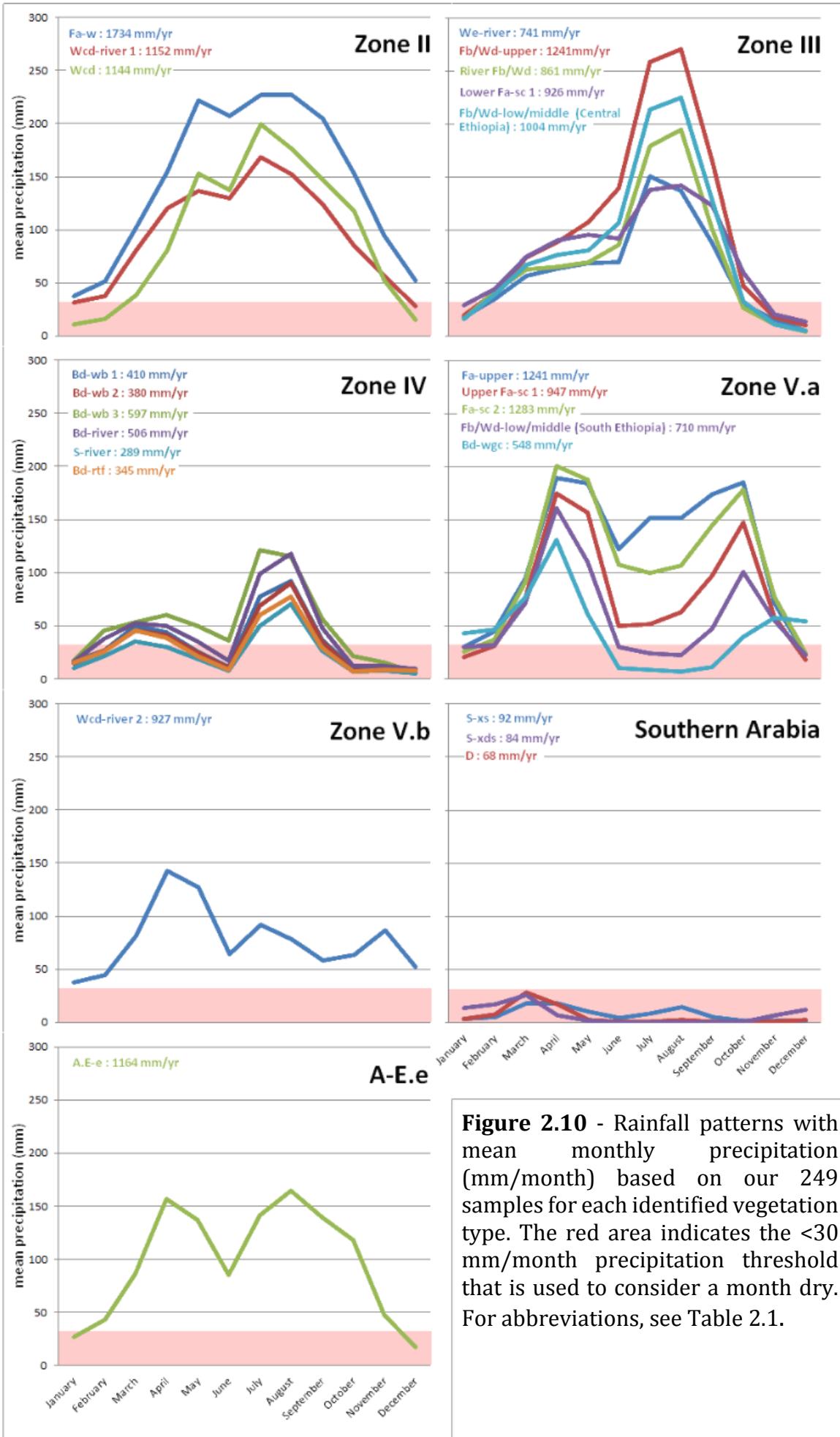


**Figure 2.9** - Map of the modern pollen samples used in this study showing the vegetation types and sub-types inferred from our new approach. Cross, diamond and star symbols distinguish soil, river and spring samples (see Table 2.1 for abbreviations). (see Table 2.1 for abbreviations).

## 2.6.4. Rainfall patterns

The final vegetation types and sub-types correspond to seven distinct rainfall patterns. These seven rainfall patterns are visible on the **Figure 2.10**. Details of mean annual rainfall and seasonal boundaries are directly visible in **Figure 2.10**. We have not included the rainfall pattern for the springs in **Figure 2.10** because the rainfall pattern is very variable. The only constant is that the rainfall patterns of Bd-spring 1 and 2 are bimodal.

The seven rainfall patterns are as follows. (1) The unimodal rainfall pattern composed of samples from Fa-w, Wcd-river 1, and Wcd-rt corresponds to the rainfall pattern in the zone II of Diro et al. (2008) (**Figure 2.2**) and is named here “zone II”. (2) The unimodal rainfall pattern composed of samples from Fa-w, lower Fa-sc 1, Wcd-river 1, Wcd-rt, and Fb/Wd-low/middle in Central Ethiopia corresponds to the rainfall pattern in the zone III of Diro et al. (2008) and is named here “zone III”. (3) The bimodal rainfall pattern composed of samples from Bd-wb 1, Bd-wb 2, Bd-wb 3, Bd-river, S-river and Bd-rtf corresponds to the rainfall pattern in the zone IV of Diro et al. (2008) and is named here “zone IV”. (4) The bimodal rainfall pattern composed of samples from Fa-upper, upper Fa-sc 1, Fa- sc 2, Fb/Wd-low/middle in Southern Ethiopia, and Bd-wgc corresponds to the main rainfall pattern in the zone V of Diro et al. (2008) and is named here “zone V.a”. (5) The rainfall pattern composed of samples from only Wcd-river 2 has no real dry season. This rainfall pattern is particular, it does not correspond to the main pattern of the zone V visible in Diro et al. (2008) but is still sub-type of the zone V (**Figure 2.2**). It is named here “zone V.b”. (6) The “rainfall pattern composed of all Southern Arabia samples (samples of D, S-xds and, S-xs) has very low precipitation without rainfall season. It is named here “Southern Arabia”. (7) The bimodal rainfall pattern composed of samples from only A.E-e has a bimodal pattern with peaks in April and August and has a dry season from December to January, and has a drop in June. It has no equivalent rainfall pattern in the zones of Diro et al. (2008) and is simply named here “A.E-e”.



**Figure 2.10** - Rainfall patterns with mean monthly precipitation (mm/month) based on our 249 samples for each identified vegetation type. The red area indicates the <30 mm/month precipitation threshold that is used to consider a month dry. For abbreviations, see Table 2.1.

It is therefore clearly possible to assign main types of rainfall patterns to the vegetation that have inferred from the pollens. This shows that the pollen spectra are able to capture differences in the floristic composition of the present-day vegetation that can be explained by differences in the rainfall regime. By extent, they may be used on the fossil record to infer the unimodality or bimodality of annual precipitations as well as the length and periods of the dry and wet seasons. For the sake of simplicity and clearness, we have grouped the patterns into 7 sets, but it is possible to see differences within these sets. For example, we note that Bd-wb 3 has a main rainfall season that is less marked than Bd-wd 1, so within Bd subtle differences in the rainfall pattern are marked in the pollen signal. But, in some cases, however, this is not applicable. For example, Fb/Wd- low/middle vegetation types include samples from southern Ethiopia with typical zone V.a rainfall pattern and samples from central Ethiopia with typical zone III rainfall pattern. It is surprising that similar pollen signal (Fb/Wd-low/middle) is obtained from vegetation types occurring under different rainfall patterns. Although the Fb/Wd-low/middle samples form a coherent whole, there are slight differences in the plants present but these are not clearly seen in the pollen spectra. It is therefore important to take it into account for future inferences via pollen.

### **2.6.5. Morphologies and vegetation strip of rivers**

This study allowed to differentiate several azonal and local vegetation types. Among the 24 final vegetation type and sub-types, the river ones (S-river, Bd-river, Bd-rtf, We-river, Fb/Wd-river, Wcd-river 1, and Wcd-river2) are particularly interesting because they allow to associate the pollen with the river morphology (meandering or straight) and type of vegetation strip along the river banks (notably the width) (**Table 2.2**). It is for example possible to say that a Bd-river pollen-inferred vegetation implies the presence of a meandriform and perennial river with a 50-100m open forest vegetative strip along its banks. Details are given for the seven pollen-inferred vegetation **Table 2.2**.

Post-hoc inferred vegetation	River morphologies <sup>a</sup>	River types <sup>a</sup>	Vegetation strip of the river
S-river	meandering	perennial or seasonal	200 m wide strip <sup>a,5</sup> of <i>Tamarix</i> - <i>Boscia</i> very open woodland/shrubland along river banks <sup>4</sup>
Bd-river	meandering	perennial	50-100m wide strip <sup>a,5</sup> of open <i>Acacia</i> - <i>Tamarix</i> forest flooded at the rainy season along the river banks <sup>4,2</sup>
Bd-rtf	very meandering	perennial	100m wide strip <sup>a,5</sup> of <i>Tamarix</i> forest flooded at the rainy season which occur between oxbow lake and along river banks followed by 200-400 m wide strip of open forest <sup>1,4</sup>
We-river	straight	perennial	50 to 100m wide strips <sup>a,5</sup> of <i>Acacia</i> woodland along river banks <sup>4</sup>
Fb/Wd-river	straight	perennial	100 to 300m wide strips <sup>a,5</sup> of <i>Ficus</i> / <i>Phoenix abyssinica</i> forest along river banks <sup>2,4</sup>
Wcd-river 1	straight	perennial	250-500 m wide strips <sup>a,5</sup> along <i>Acacia</i> / <i>Phoenix reclinata</i> forest along river banks <sup>1,2</sup>
Wcd-river 2	meandering	perennial	3 km wide strip <sup>a,5</sup> of alternation <i>Acacia</i> - <i>Cordia</i> forest, <i>Acacia</i> - <i>Cordia</i> woodland and <i>Acacia</i> - <i>Cordia</i> open woodland in function of soil humidity along river banks <sup>3,4</sup>

**Table 2.2** - Relationship between vegetation inferred by pollen and the characteristics of rivers. a : Personal observation using Google Earth images (1984 to 2020); 1 : Bonnefille et al. (1987); 2 : Buchet (1982); 3 : Carr (1998); 4 : Friis et al. (2010); 5 : Kobayashi et al. (2017). For abbreviations, see Table 2.1.

### 2.6.6. Most characteristic pollen types

At the level of the 24 final vegetation types and sub-types, if we look at the 10 most characteristic pollen types of each of these vegetation (**Table 2.3**) on the base of T-values (**Annexe table 2.6**), we note that arboreal pollen types are more often used than non-arboreal pollen types for vegetation characterisation. This is in line with traditional palynological analyses. For the most open and xerophytic vegetation (S-river, S-xs, S-xds, D), however, arboreal pollen types are used less often than non-arboreal pollen types. But also for A.E.e where it is more unexpected. For some vegetation types, such as Fa-w, the pollen taxa markers correspond to botanical markers identified by botanists (*Macaranga*, *Pouteria* and *Phoenix reclinata*) (Friis et al. (2010). Overall, only 10% to 40% of the pollen taxa identified by our approach as the most characteristic for each km-group correspond to botanical markers identified by botanists (**Table 2.3**). Thus, these results are not surprising as this is also the case in traditional pollen studies.

	Arboreal plants (AP)	Non-arboreal plants (NAP)	Undifferentiated (I)
S-river	<i>Shirakia</i> - type <i>elliptica</i> ; <i>Tamarix</i> ; <i>Acacia</i>	<i>Tribulus</i> ; <i>Scabiosa</i> - type; <i>Pupalia</i> - type <i>lappacea</i> ; <i>Euphorbia</i> - type <i>indica</i> ; <i>Digera</i> - type <i>muricata</i> ; <i>Aerva</i> - type <i>javanica</i>	Nyctaginaceae
Wcd-river 1	<i>Macaranga</i> - type; <i>Phoenix reclinata</i> - type; <i>Lannea</i> - type; <i>Dodonaea viscosa</i> - type; <i>Sesbania sesban</i> ; <i>Hagenia abyssinica</i> ; <i>Trilepismum</i> - type <i>madagascariensis</i>		<i>Acalypha</i> ; <i>Anthospermum</i> ; <i>Urticaceae</i> T-type
Wcd-river 2	<i>Schefflera abyssinica</i> - type; <i>Combretaceae</i> ; <i>Polyscias fulvo</i> - type; <i>Psydrax</i> - type <i>schimperiana</i>	<i>Cardiospermum halicacabum</i> - type; <i>Urticaceae</i> D-type; <i>Hermannia</i> - type; <i>Vahlia somalensis</i> - type	<i>Hibiscus</i> / <i>Pavonia</i> ; <i>Croton</i> - type
FbWd-river	<i>Olea</i> ; <i>Strophanthus</i> - type <i>mirabilis</i> ; <i>Craterispermum schweinfurthii</i> - type; <i>Eucalyptus</i> - type	<i>Vicia</i> ; <i>Commelina</i> - type <i>forskalei</i> ; <i>Canavalia</i> - type	<i>Dyschoriste</i> - type <i>radicans</i> ; <i>Crotalaria</i> ; <i>Diplolophium</i> - type
We-river	<i>Canthium</i> - type; <i>Celastraceae/Hippocrateaceae</i> ; <i>Capparidaceae</i> ; <i>Caucanthus</i> - type	<i>Commelina</i> - type <i>benghalensis</i> ; <i>Justicia anselliana</i> - type; <i>Polygonum senegalense</i> - type ; <i>Lamiaceae</i>	<i>Hypoestes</i> - type; <i>Acanthaceae</i>
Bd-river	<i>Suaeda monoica</i> - type; <i>Tapinanthus</i> - type <i>globiferus</i> ; <i>Ximenia</i> ; <i>Justicia</i> - type <i>odora</i>	<i>Paronychia</i> - type; <i>Cissampelos</i> - type <i>mucronata</i> ; <i>Celosia</i> - type <i>trigyna</i> ; <i>Corchorus</i> - type	<i>Acanthus</i> - type; <i>Rumex</i>
D		<i>Cleome</i> - type <i>brachycarpa</i> ; <i>Cyperaceae</i> ; <i>Plantago</i>	<i>Limeum</i> ; <i>Cissus</i> ; <i>Fagonia</i> ; <i>Rumex</i> ; <i>Amaranthaceae/Chenopodiaceae/Caryophyllaceae</i> ; <i>Lamiaceae</i> ; <i>Brassicaceae</i> ; <i>Tubuliflorae</i> ; <i>Apiaceae</i>
S-xds	<i>Dodonaea viscosa</i> - type	<i>Plantago</i> ; <i>Heliotropium steudneri</i> - type	<i>Armeria</i> / <i>Limonium</i> ; <i>Amaranthaceae/Chenopodiaceae/Caryophyllaceae</i> ; <i>Brassicaceae</i> ; <i>Cichorieae</i> ; <i>Tubuliflorae</i> ; <i>Apiaceae</i>
S-xs	<i>Piliostigma</i> ; <i>Pinus</i>	<i>Cleome</i> - type <i>brachycarpa</i> ; <i>Aerva</i> - type <i>lanata</i> ; <i>Zygophyllum simplex</i> - type; <i>Lotus arabicus</i> - type	<i>Justicia</i> - type; <i>Brassicaceae</i> ; <i>Zygophyllaceae</i> ; <i>Resedaceae</i>
Bd-spring 1	<i>Hyphaene</i> - type; <i>Tapura fischeri</i> - type; <i>Suaeda monoica</i> - type; <i>Dorstenia</i> ; <i>Duosperma</i>	<i>Typha</i> ; <i>Pupalia</i> - type <i>lappacea</i> ; <i>Heliotropium steudneri</i> - type	<i>Amaranthus</i> ; <i>Capparidaceae</i>
Bd-spring 2	<i>Hyphaene</i> - type; <i>Adansonia</i> ; <i>Aeschynomene baumii</i> - type; <i>Glenniea</i> - type <i>africana</i> ; <i>Ximenia</i>	<i>Cyperaceae</i> ; <i>Echium</i> - type; <i>Gunnera perpensa</i> ; <i>Onagraceae</i>	<i>Acanthaceae</i>
Bd-wb 1	<i>Boscia</i> - type; <i>Hildebrandtia</i> - type <i>obcordata</i> ; <i>Capparis</i> ; <i>Loranthaceae</i> ; <i>Ornocarpum</i> - type <i>trichocarpum</i> ; <i>Acacia</i> ; <i>Crateva</i> - type <i>adansonii</i>	<i>Euphorbia</i> - type <i>indica</i> ; <i>Evolvulus</i> - type	<i>Acanthus</i> - type
Bd-wb 2	<i>Suaeda monoica</i> - type	<i>Digera</i> - type <i>muricata</i> ; <i>Aerva</i> - type <i>javanica</i> ; <i>Tribulus</i> ; <i>Ocimum</i> - type <i>basilicum</i> ; <i>Tacazzea</i> - type <i>apiculata</i> ; <i>Euphorbia</i> - type <i>indica</i> ; <i>Kohautia caespitosa</i> - type	<i>Nyctaginaceae</i> ; <i>Abutilon</i>
Bd-wb 3	<i>Flueggea virosa</i> - type; <i>Grewia</i> - type; <i>Acacia</i> ; <i>Cordia</i> ; <i>Suaeda monoica</i> - type	<i>Amaranthus</i> - type; <i>Psilotrichum</i> - type <i>elliottii</i> ; <i>Hyptis pectinata</i> - type	<i>Tubuliflorae</i> ; <i>Abutilon</i>
Bd-rtf	<i>Tamarix</i> ; <i>Phyllanthus</i> - type <i>reticulatus</i> ; <i>Phyllanthus</i> - type <i>muellerianus</i> ; <i>Phyllanthus</i> - type <i>englerti</i>	<i>Cocculus</i> - type; <i>Andrachne aspera</i> - type; <i>Cleome</i> - type <i>parvipetala</i> ; <i>Phyllanthus</i> - type <i>boehmii</i> ; <i>Cleome</i> - type <i>orthocantha</i>	<i>Chenopodiaceae</i>
Bd-wgc	<i>Annona</i> - type <i>senegalensis</i> ; <i>Lannea</i> - type; <i>Rauvolfia</i> ; <i>Euphorbia tirucalli</i> - type; <i>Cussonia</i> - type; <i>Trichilia</i> - type <i>emetica</i> ; <i>Bridelia</i> - type; <i>Argomuellera</i> - type <i>macrophylla</i>	<i>Cyathula</i> - type <i>orthocantha</i> ; <i>Justicia anselliana</i> - type	
Wcd	<i>Anogeissus</i> - type <i>leiocarpa</i> ; <i>Entada</i> - type; <i>Sterculia</i> - type; <i>Pterocarpus</i> - type <i>lucens</i> ; <i>Combretum</i> - type <i>aculeatum</i> ; <i>Combretaceae</i> ; <i>Lannea</i> - type	<i>Commelina</i> - type <i>forskalei</i> ; <i>Justicia striata</i> - type	<i>Gentianaceae</i>
FbWd-low/middle	<i>Dodonaea viscosa</i> - type; <i>Olea europaea</i> - type; <i>Acokanthera</i> - type <i>schimperi</i> ; <i>Embelia</i> - type; <i>Keetia</i> - type <i>gueinzii</i> ; <i>Milletta</i> - type <i>ferruginea</i> ; <i>Clematis</i> - type; <i>Myrsine africana</i>		<i>Clutia</i> ; <i>Vitaceae</i>
FbWd-upper	<i>Myrica</i> ; <i>Ericaceae</i> ; <i>Protea</i> - type; <i>Juniperus</i> - type <i>procera</i> ; <i>Olea</i> ; <i>Ebenaceae</i> ; <i>Myrsine africana</i>	<i>Plantago</i>	<i>Anthospermum</i> ; <i>Hypericum</i>
Fa-sc 1	<i>Podocarpus</i> ; <i>Olea capensis</i> - type; <i>Celtis</i> ; <i>Rhamnaceae</i> ; <i>Apocynaceae</i> ; <i>Zantha golungensis</i> - type; <i>Alangium chinense</i> ; <i>Warburgia</i> - type	<i>Swertia</i> - type <i>kilimandscharica</i>	<i>Pterodiscus</i>
Fa-sc 2	<i>Eugenia/Syzygium</i> ; <i>Prunus africana</i> - type; <i>Apodytes dimidiata</i> ; <i>Olea capensis</i> - type; <i>Zanthoxylum/Fagaropsis</i> ; <i>Macaranga</i> - type; <i>Sideroxylon</i> - type <i>oxyacantha</i> ;	<i>Tiliacora</i> - type <i>funifera</i> ; <i>Sebaea brachyphyllea</i> - type	
Fa-upper	<i>Hagenia abyssinica</i> ; <i>Schefflera volkensii</i> - type; <i>Rubus pinnatus</i> - type; <i>Rapanea melanophloeos</i> - type; <i>Celastraceae/Hippocrateaceae</i> ; <i>Galiniera saxifraga</i> ; <i>Schefflera abyssinica</i> - type; <i>Prunus africana</i> - type	<i>Lobelia rhynchopetalum</i> - type	<i>Isoglossa</i>
Fa-w	<i>Cupressaceae</i> ; <i>Trilepismum</i> - type <i>madagascariensis</i> ; <i>Alchornea</i> ; <i>Macaranga</i> - type; <i>Chiornanthus mildbraedii</i> - type; <i>Pouteria</i> - type; <i>Maesa lanceolata</i> - type; <i>Phoenix reclinata</i> - type	<i>Sericostachys</i> - type <i>scandens</i>	<i>Impatiens</i>
A.E		<i>Cerastium afromontanum</i> - type; <i>Monocotyledoneae</i> ; <i>Lythrum</i> ; <i>Cyperaceae</i> ; <i>Ranunculus oreophytus</i> - type; <i>Hydrocotyle mannii</i> - type; <i>Lobelia anceps</i> - type	<i>Alchemilla</i> ; <i>Caryophyllaceae</i> ; <i>Apiaceae</i>

**Table 2.3** - The 10 most characteristic pollen types of each vegetation type inferred by the pollen based of the T-values (Annex table 2.6). Taxa are listed from the most to the least important. For abbreviations, see Table 2.1.

Characteristic pollen types for riparian and azonal vegetation types were easily identified as *Tamarix* for Bd-rft and *Hyphaene* for Bd-spring 1 et 2 in agreement with the abundance of these trees in riparian *Tamarix* forests and groundwater-fed woodlands. For dynamic depositional environments like rivers, however, correlations with the surrounding vegetation are not obvious, which is not surprising since the pollen spectra integrate pollen grains likely coming from vegetation much further upstream. Yet, six different river vegetation types were identified by our approach, which shows that riparian formations carry distinct pollen signals according to the vegetation zones crossed by the river .

## 2.7. Discussion

### 2.7.1. Usefulness for environmental reconstructions

24 final vegetation types and sub-types can be inferred from pollen assemblages using our clustering approach, combined with t-value calculation (**Table 2.1**). Using our new approach of pollen score calculation, based on the combination of V and V-river score, we have seen that it is possible to attribute specific values to each of these post-hoc inferred vegetation and we have determined significance intervals for these specific values (**Figure 2.5**). This approach allows a rigorous discussion of the affiliation of a pollen spectrum to one of these 24 post-hoc inferred vegetation. These 24 final vegetation types and sub-types inferred are distributed among eight regional vegetation and 22 are sub-types of these regional vegetation. This is the first time that pollen scores have been successfully developed to reconstruct vegetation with a spatial resolution lower than the regional scale. In East Africa, reconstructions using score calculations have been limited to the supra-regional or regional scale for Ethiopia, Kenya, Tanzania, and Uganda (Bonnefille et al., 2004; Vincens et al., 2006b). This is also the case for Central and West Africa (Hély et al., 2006; Izumi and Lézine, 2016; Lebamba et al., 2009).

Within these 22 sub-types, nine are local and azonal vegetation fed by rivers or springs. The difficulty in the separation of the local and regional pollen signal has long been known (Bonnefille, 1984) and had not yet been done by score calculation methods. Riparian environments, in most palynological studies are usually ignored (Bonnefille et

al., 2004; Hély et al., 2006; Izumi and Lézine, 2016; Lebamba et al., 2009; Vincens et al., 2006b), and aquatic taxa are usually removed from the analyses as originally proposed by Jolly et al. (1998). Our results therefore constitute a major step forward for future application in the study of Hominini sites in East Africa, because they are very frequently associated with riparian or source water contexts (Barboni et al., 2019; Campisano and Feibel, 2008; Levin, 2015). The reconstructions of these local riparian environments where the Hominini fossils are located, are not clearly defined yet by palynological or faunal studies, which normally define them as "swamp", "wetland" or "floodplain" (Bonnefille et al., 1987; Bonnefille, 2010; Bonnefille et al., 2004; Reed, 2008).

Our results show that the pollen type *Hyphaene*-type is a main marker of spring as suggested by Barboni et al. (2019). This is an important step because it is supposed that springs have an important role in the hominin evolution in East Africa (Ashley et al., 2009; Cuthbert et al., 2017; Cuthbert and Ashley, 2014). It is interesting to note that there is no clear distinction between Ethiopian and Tanzanian springs despite the regional differences in vegetation and groundwater (alkaline water in Ethiopia and fresh water in Tanzania: personal observation). The important parameter to differentiate the spring vegetation is their distance from the spring head.

We have not been able to separate the Afroalpine vegetation (A) and Ericaceous belt (E) pollen signal. The samples of these two vegetation are in the A.E-e vegetation subtype. An explanation is that all samples of Afroalpine vegetation (A) are in the Sanetti plateau (Ethiopia) and correspond to a particular area with isolated *Erica* shrubs (Mohammed et al., 2007). Miehe and Miehe (1994) suggested that this area represents the potential extent of *Erica*-dominated vegetation in the absence of natural or anthropic fires, hence our difficulty in the separation of Afroalpine vegetation (A) and Ericaceous belt (E) samples. Another difficulty is at the lower limit of the Afromontane forest (Fa) where Friis et al. (2010) observe "moist evergreen Afromontane forest, bushland, woodland and wooded grassland". Although several samples are available in this area, our analyses did not reveal any specific pollen composition for this. We don't know if this is due to the absence of a specific pollen signal or to a limit of our approach.

For Fa-w, Bonnefille et al. (1993) indicate that the Fa-w samples collected near Gore (western Ethiopia) come from the Transitional rain forest (Fe) type. However, Friis

et al. (2010) describe the vegetation at the location of the sampling sites of Fa-w samples (near Gore) as "Primary or mature secondary moist evergreen afromontane forest", a sub-type of Fa type. Normally, the elevation of Fa-w samples (on average 1600 m a.s.l.) is too low for Fa (low limit of Fa: 1800 m a.s.l.). But in this part of Western Ethiopia, the "Primary or mature secondary moist evergreen afromontane forest" has the particularity of occurring below the calculated lower elevation limit of the Fa, and grows at the same altitude as the Transitional rain forest (Fe) (Friis et al., 2010). In addition, the present plants described by Bonnefille et al. (1993) corresponds to vegetation of "Primary or mature secondary moist evergreen afromontane forest" (Friis et al., 2010). This is why we preferred to define Fa-w as Secondary moist evergreen afromontane forest rather than as Transitional rainforest.

Pollen assemblages are able to discriminate seven riparian vegetation types, which relate to distinct floristic composition. A closer look at the data also shows that distinct floristic composition of riparian forest is not only linked to the surrounding regional vegetation (Friis et al., 2010a), but also to river morphology (Carr, 1998). Carr (1998) showed that vegetation patterns change along the Omo River (southwestern Ethiopia) as a function of channel structure. These vegetation changes are also found in the studied pollen assemblage, this is the case for the Wcd type where Wcd-river 1 and Wcd-river 2 sub-types are associated with a straight and meandering river respectively. However, it is not possible for us to track fine changes in channel structure as Carr (1998) did, we can only estimate the main types of channel structure. Similarly, it is possible to associate pollen assemblages with the width of the vegetative strip along rivers (**Table 2.2**). Again, this agrees with the observations of Carr (1998) but also with our satellite observations. The structure of the river and the width of the vegetation strip are important elements because they will influence the fauna present.

This is the first time that so many climatic and soil parameters have been compared with pollen data for Eastern Africa. We found that the parameter that best explained our these 24 post-hoc inferred vegetation was the percentage of tree cover. It is now possible to quantify the percentage of tree cover based on pollen data. To our knowledge, this has never been done before. Our approach also allows inferring the mean annual temperature and the mean annual precipitations. Indeed, these two climatic parameters are the next best variables which explain the pollen-inferred vegetation types.

The mean temperature of warmest quarter and the mean precipitation of driest quarter also explain the inferred vegetation. They appear as important limiting parameters of the vegetation. The seasonality of precipitation and the amplitude of alpha are also in a good position as observed in other parts of the world (Barboni and Bonnefille, 2001). Bonnefille et al., (2004) used the method of Best Analogs (Bonnefille et al., 1992) and the method of Plant Functional Type (Peyron et al., 2000) to estimate temperature and precipitation. The resolution of these parameters via the extrapolation of our inferred vegetation is comparable to the resolution of the Best Analogs method and is better than the resolution of the Plant Functional Type method. For the soil characteristics, they are far less important than climate in determining floristic composition and the presence or absence of plant species. The NPP remains moderately interesting. The NPP is related to the distribution of vegetation types but not sub-types. This is an astonishing result but can be explained by the small differences in NPP between neighbouring vegetation sub-types.

Finally, we have seen that pollen allows with our approach to make consistent estimates of rainfall patterns for Ethiopia. This is also a novelty for Eastern Africa. The rainfall pattern estimated with our approach will allow a better understanding of the temporal variation of the Intertropical Convergence Zone (ITCZ) in Ethiopia (Diro et al., 2011).

Our study shows that pollen data in presence-absence provide better vegetation inference than pollen data in abundance scale for pollen samples obtained from dynamic depositional environments, such as river beds and margins (**Annexe figure 2.7**). Therefore, fossil pollen samples from Hominini sites, which are frequently found in fluvial contexts (eg., Saylor et al., 2019), will need to be treated differently than paleosoil samples. The availability of satellite products has made possible to develop a pollen-tree cover response function. Pollen-inferred tree cover will be very interesting to compare with paleo-tree cover estimates based on  $\delta^{13}\text{C}$  at hominin sites (eg., Cerling et al., 2011).

## 2.7.2. A more objective approach

Some authors such as Izumi and Lézine (2016) used thresholds for Poaceae that can be compared to our abundance classes, but these thresholds are defined subjectively for each pollen type whilst our abundance scale is based on statistical analyses. Similarly, the choice of assigning a pollen type to a type of vegetation remains arbitrary or according to the vegetation observed by botanists. However, the statistical analyses of this study (**Table 2.3**) show that the most characteristic pollen types of vegetation are not always those associated with the plants used by botanists to describe the vegetation. We have shown here that it is possible to solve these problems of subjectivity for the thresholds and the attribution of vegetation types by using the T-values of the pollen types to define the weight of each pollen type for each vegetation. Still in this idea of objectivity, the significance intervals mentioned above are an important addition to the previous methods of calculating pollen score, which are based only on the vegetation with the highest score without indicating whether the value is significant or not (Bonnefille et al., 2004).

In contrast with previously pollen study in Eastern Africa for modern and fossil pollen samples (Bonnefille, 2010), we do not work in pollen percentages, because we may observe variation of pollen percentage for the same vegetation. These variations may depend on the sampling area, the sample preparation and/or the pollen analyst. We prefer to use the scale of abundance commonly used in modern ecological studies (Borcard et al., 2018) to minimize the importance of these variations. The method of abundance classes has the advantages of presence/absence and percentage method, without the disadvantages. It minimizes the problems of over-representation (eg., Cyperaceae, Poaceae) or under-representation (eg., arboreal pollen) of pollen types. Our classes of abundance have the same role as the square root (Prentice et al., 1996; Vincens et al., 2006b), but are more relevant because they are specifically calculated for the pollen counts of Eastern Africa.

To reduce the noise due to occasional long-transport pollen grains, the current value used in East Africa is 0.5% and it is based on Jolly et al. (1998) and Prentice and Webb III (1998). Nevertheless, these two studies only repeat the value of Prentice et al. (1996) used for Europe without justification for transposition to Africa, which is a

subjective choice of the authors. Based on our graphical analyses (**Annexe figure 2.2**), we believe that a threshold of 0.2% is more appropriate for Africa. Izumi and Lézine (2016) also suggest this with empirical observations. This changes a lot for pollen score calculations in Africa. For our study, for example, using a threshold of 0.2% removes only 9.5% of total pollen occurrences compared to 43.5% for a threshold of 0.5% (**Annexe figure 2.2**).

## 2.8. Conclusions

K-means clustering and machine learning applied to a modern data set of 283 pollen spectra from northeastern Africa allows inferring 24 distinct vegetation types and sub-types, objectively. These 24 types and sub-types are well-constrained floristically and climatically, and are characterized by specific tree cover ranges. Based on a thorough statistical analysis of the data, most of the pollen taxa and counts were considered, and the floristic and climatic signal carried out by aquatic taxa was fully exploited. Hence, our new approach allows identifying nine riparian vegetation sub-types (two of springs and seven of rivers) which not only relate to regional vegetation composition, climate, but also to the type of wet edaphic conditions. Our study shows that riparian vegetation types in north part of Eastern African Rift carry distinct pollen and climatic signals, which was largely ignored before. Hence, at fossil Hominini sites in Ethiopia that are often found linked to riparian or spring environments, this new approach is likely to provide more informative vegetation inferences than before. Also, this new approach proposes to treat pollen data differently according to the depositional context they originate. This is the case for pollen spectra from dynamic depositional environments which should be considered in terms of presence/absence rather than percentages or even abundance classes. Finally, our new approach takes full advantage of the multifactorial nature of the pollen spectrum. It allows, in addition to the identification of vegetation, the extrapolation of subtle information such as precipitation regimes, temperature seasonality, percentage of tree cover, river morphology (for riverine vegetation), etc. This permits, in the future, to go much further in the analyses than other proxies such as phytoliths and  $\delta^{13}\text{C}$ .

These results pave the way for re-discuss the reconstructions of the habitats of early hominids with greater finesse and to provide more environmental information.

## **2.9. Acknowledgments**

This work was carried out as part of BB's PhD (PhD fellowship financed by equity of the University of Aix-Marseille). BB's research work was supported by the Agence Nationale de la Recherche (HADoC Project to DB, ANR-17-CE31-0010-02). Fieldwork in Ethiopia was made possible thanks to the financial support of CEREGE (APIC-2017), the NSF-RHOI project (2009), and the logistic support of CFEE at Addis Abeba. Fieldwork would not have been possible without the precious help of Maurice Taieb and Shoki Alisaïd for visas, the collaboration with Dr Teshome Soromessa, Dr Wonde Girmay and Melaku Wondafrash at Addis Ababa University, and the wise advice of Raymonde Bonnefille in the field in 2009, Berhane Asfaw, Elema Hamed, Dr Admasu Adi, and our driver Amare.

### **3. *Australopithecus afarensis* relié aux environnements boisés : reconstruction palynologique à haute résolution spatiale pour les paléo-végétations des Formations de Hadar et Busidima (Afar, Ethiopie)**

• Titre de l'article : *Australopithecus afarensis* linked to woodlands and forests in Northern Awash Basin: palynological reconstruction with high spatial resolution for the paleovegetation of Hadar and Busidima Formations (Afar, Ethiopia)

• Auteurs : Benjamin Bourel<sup>1</sup>, Doris Barbóni<sup>1</sup>, Chris J. Campisano<sup>2</sup>, Andrew S. Cohen<sup>3</sup>

<sup>1</sup>CEREGE, Aix Marseille Université, CNRS, IRD, INRAE, Coll. France, Technopole Arbois, 13545 Aix en Provence cedex 4, France

<sup>2</sup>Institute of Human Origins, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA

<sup>3</sup>Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA

• Revue : Version préliminaire qui sera modifiée avant sa soumission, peut-être pour *Journal of Human Evolution*, en discussion avec les co-auteurs.

• Résumé : The Pliocene Hadar and Busidima Formations (Afar, Ethiopia) revealed rich fossil faunas often associated with hominin remains (*Australopithecus* spp. and early *Homo*) between 2.35 and 4.00 million years (Ma). 42 fossil pollen samples related to these two formations, including 21% new samples from the HSPDP cores NAW-1A, are for the first time interpreted in terms of paleovegetation at the regional scale and at the local scale, through a new method based on machine learning. Our results show that 1) significant changes in the rainfall pattern involving major contraction of the Intertropical Convergence Zone (ITCZ) occurred at Hadar at  $3.42 \pm 0.03$  Ma,  $3.39 \pm 0.03$  Ma,  $3.357 \pm 0.02$  Ma, ca. 3.28 Ma, ca. 3.19-3.20 Ma, and ca. 2.53-2.58 Ma; 2) the global cooling episodes related to the Marine Isotope Stage (MIS) MG4, MG2, and M2 had significant impact on the regional and local vegetation at Hadar, and likely resulted in a temperature decrease of ca.  $-6.5^{\circ}\text{C}$  to ca.  $-10.5^{\circ}\text{C}$  compared to the present; 3) several samples indicate that azonal vegetation types such as riverine and spring forests were frequent at Hadar and Dikika, and likely offered strikingly more wooded habitats than the surrounding open and dry landscape. This study shows that despite important variations of vegetation at regional

scale during the occurrence of *Australopithecus afarensis* in the Northern Awash Basin, vegetation at local scale is still forest or woodland thanks to the river and springs. This shows that the ecology of *Australopithecus afarensis* may not be as varied but rather associated with closed or semi-open environments (at least for *Australopithecus afarensis* in the Northern Awash Basin).

- Mots clefs : pollen, Hominini, paleoenvironment, riparian vegetation, microhabitat

### 3.1. Introduction

The East African Rift is famous in its eastern branch (Ethiopia, Kenya and Northern Tanzania) for its great importance in the fossil record of the Hominini (Levin, 2015). The Hominini have been present in East Africa for 6 Ma with *Orrorin tugenensis* dated at ca. 6.0 Ma to 5.7 Ma (Tugen Hills, Kenya) followed by *Ardipithecus kadabba* dated at ca. 5.8 Ma to 5.2 Ma (Middle Awash Valley, Ethiopia) and *Ardipithecus ramidus* dated at ca. 4.4 Ma (Middle Awash Valley, Ethiopia) (Haile-Selassie et al., 2004; White et al., 2009, 2006). Speciation events in this geographical area have setup the first known *Australopithecus* species with *Australopithecus anamensis* at ca. 4.2 Ma to 3.6 Ma (Ethiopia and Kenya) (Haile-Selassie et al., 2010b; White et al., 2006). From this, an important diversification of Hominini began to take place in this part of the rift with the development of new *Australopithecus* species and the diversification of new genera with *Paranthropus* (first occurrence: ca. 2.7 Ma) and later *Homo* (first occurrence: ca. 2.8 Ma) (Behrensmeyer and Reed, 2013; Campisano et al., 2017).

Among the various major paleoanthropological research basins in Eastern Africa, the Northern Awash Basin, also called the Lower Awash Valley (Ethiopia: Pliocene and Lower Pleistocene) (**Figure 3.1**) is particularly rich in fossils (Behrensmeyer and Reed, 2013; Campisano, 2019). It includes several world-famous research areas, such as 1) Hadar (main period studied: 2.95 Ma to 3.45 Ma) (Campisano and Feibel, 2008) with the discovery of specimen AL 288-1, better known under the nickname Lucy (Johanson et al., 1978; Johanson and Taieb, 1976); 2) the Gona research area (main period studied: ca. 3.4 Ma to 1.6 Ma) (Quade et al., 2004) which is rich in stone tools with old Oldowan tools (ca. 2.6 Ma) (Semaw et al., 2003); 3) the Dikika area (main period studied: ca. 4.0 Ma or 3.8 Ma to 2.7 Ma: Pliocene) (Wynn et al., 2008) with the famous juvenile of early hominin

skeleton DIK-1/1 (3.3 Ma) nicknamed Selam (Alemseged et al., 2006) and the oldest evidence of stone tool use (3.39 Ma) (McPherron et al., 2010); 4) the “upper” Pliocene and “lower” Pleistocene area of Ledi-Geraru (southern and central sub-area: ca. 3.4-2.9 Ma; eastern sub-area: 2.4-3.0 Ma) (Campisano, 2019) known for its record of the oldest fossil of *Homo* (LD 350-1, left hemi-mandible: ca. 2.8 Ma) (Villmoare et al., 2015); 5) the Burtele research area with the “Burtele foot” (BRT-VP-2/73: 3.4 Ma) associated with a new unknown arboreal Hominini species (Haile-Selassie et al., 2012); 6) the research area of Woranso-Mille (4.0-3.2 Ma) (Campisano, 2019) with transitional specimens of *Australopithecus anamensis* from 3.8 Ma to 3.6 Ma (Haile-Selassie et al., 2010b). These six research areas are within an area of only 40 km (East-West) by 60 km (North-South), and are linked to the Hadar and Busidima Formations. This makes the Northern Awash Basin a major site for the study of the Plio-Pleistocene Hominini for the study of morphological evolutions and development of stone tool cultures.

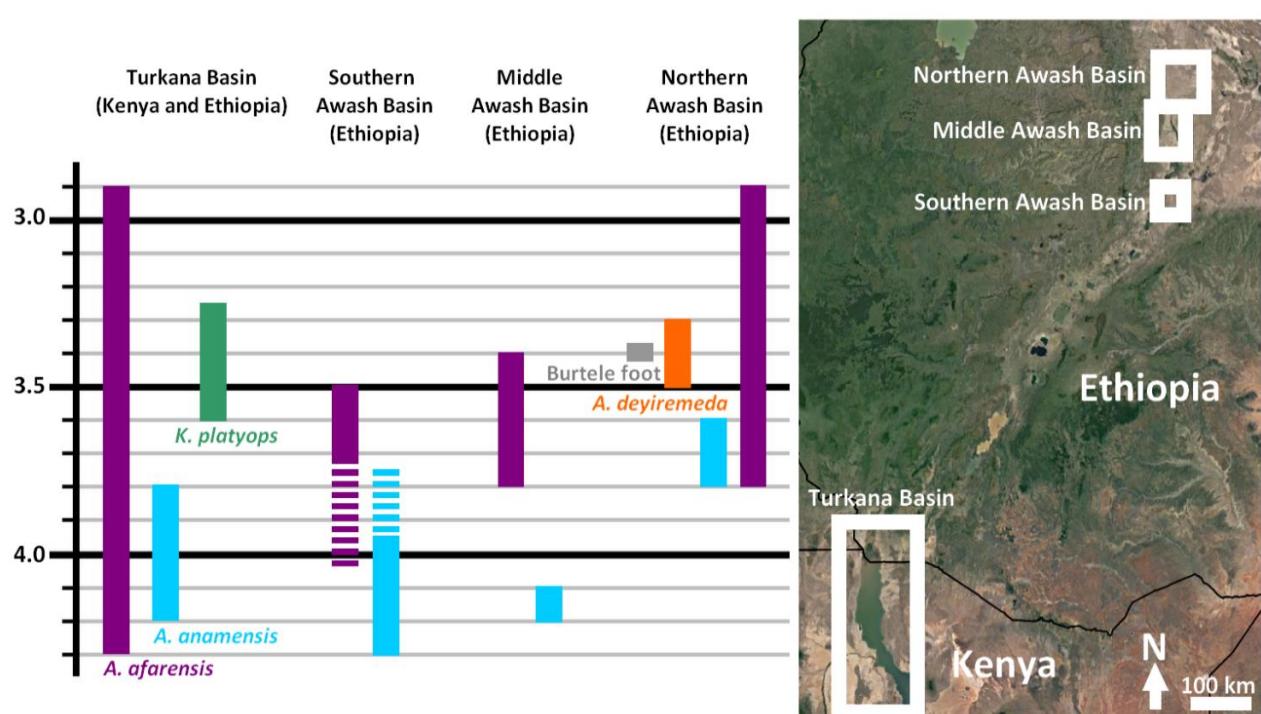
In this area, palynological studies of Pliocene strata were undertaken as early as the 1970s by Raymonde Bonnefille, whose team has provided 29 pollen spectra between ca. 4.0 Ma and 2.35 Ma (between 3.37 Ma and 3.18 Ma for 25 spectra) from Hadar (26 spectra), Gona (1 spectrum) and Dikika (2 spectra) (Bonnefille et al., 1987; Bonnefille, 2010; Bonnefille et al., 2004). Other teams provided three pollen spectra from Gona at ca. 2.6 Ma (López-Sáez and Domínguez-Rodrigo, 2009) and one from Woranso-Mille at ca. 4.0 Ma (Saylor et al., 2019). Paleoenvironmental reconstructions with good temporal resolution were proposed between 2.9 Ma and 3.8 Ma for the area at the Hadar, Gona and Dikika sites using classical palynological approaches (Bonnefille et al., 1987) and pollen score approaches based on Best Analog (BA) or Plant Functional Types (PFT) methods (Bonnefille et al., 2004) (**Figure 3.1**). These palynological studies as well as faunal analyses (Reed, 2008) indicated that during the occurrence period of *Australopithecus afarensis*, the region of Hadar, Gona and Dikika underwent large regional environmental changes with biomes ranging from humid closed to dry open vegetation. *Australopithecus afarensis* was therefore interpreted as a eurytopic species (Bonnefille et al., 2004; Reed, 2008).



**Figure 3.1** - Pollen sampling areas by locality in the Northern Awash Basin (Ethiopia).

During this period of time (2.9 Ma to 3.8 Ma), *Australopithecus afarensis* cohabited in the Northern Awash Bassin (Ethiopia) with an undetermined Hominini species that was morphologically different from the Hominini species already known at 3.4 Ma (BRT-VP-2/73: "Burtele foot") (Haile-Selassie et al., 2012), and two other *Australopithecus* species, including a transitional sub-species of *Australopithecus anamensis* from 3.8 Ma to 3.6 Ma (Haile-Selassie et al., 2010b) and *Australopithecus deyiremeda* around 3.3-3.5 Ma (Campisano et al., 2017). *Australopithecus afarensis* was also contemporary with

*Kenyanthropus platyops* which was present in the Turkana basin to the south of the Awash basin between 3.6 Ma and 3.25 Ma (Campisano et al., 2017) (**Figure 3.2**). The recent geomorphometric analyses of the maxillae confirm that *Australopithecus deyiremeda* and *Kenyanthropus platyops* were distinct species of *Australopithecus afarensis* (Spoor et al., 2016). Knowing this, without the presence of distinct habitats or ecological preferences, it is difficult to conceive this diversity of Hominini at this period and on this little geographical area (Behrensmeyer and Reed, 2013). We know today that Hominini sites frequently occur at level of river, lake or spring depositional environments (Barboni et al., 2019; Levin, 2015). This is particularly true in Hadar, as shown by the sedimentology (Campisano and Feibel, 2008). In these types of edaphic contexts, there is often the establishment of local and azonal vegetation that differs from the surrounding regional and zonal vegetation (Friis et al., 2010a; Geheb and Abebe, 2003). This is potentially the case at the level of Hominini sites in the Northern Awash Valley. Previous environmental reconstructions based on pollen or fauna may have missed these spatially very localized azonal habitats because of the low spatial resolution of their reconstructions (regional or supra-regional scale) (Behrensmeyer and Reed, 2013).



**Figure 3.2** - Chronological and spatial distribution of Hominini species in the northern part of the Eastern African Rift (Ethiopia and Northern Kenya) between 4.3 Ma and 2.9 based on Bobe et al. (2020), Behrensmeyer et al. (2013), Campisano et al. (2017) and Haile-Selassie et al. (2012).

In palynology, it is difficult to differentiate the local pollen signal from the regional signal, whether by the PFT method, BA method (Bonnefille et al., 2004) or classical approaches (Bonnefille et al., 1987). Recently, a new approach for calculating pollen scores has developed from a modern pollen data set obtained from 283 surface samples (soils and muds, mostly from Ethiopia) to solve this problem (**Part 1 of this thesis**). This new approach estimates the local paleovegetation and the surrounding regional paleovegetation. It allows identifying local vegetation types which are azonal, that is not in equilibrium with climate, such as palm woodlands which develop at groundwater-fed sites. This new approach can also be used to quantify climatic variables and tree cover based on pollen assemblages.

Hence, in order to provide new paleovegetation and paleoclimatic inferences for the Pliocene environment and potential habitats at Hadar and surrounding areas, we have applied our new objective statistical approach to 42 pollen samples from the Northern Awash Basin, including 21% new samples from the Hominin Sites Paleolakes Drilling Project (HSPDP) (Cohen et al., 2016). Using this new approach, our study aims to see whether *Australopithecus afarensis* is indeed a ubiquitous species as proposed by environmental reconstructions on a regional scale (Bonnefille et al., 2004; Reed, 2008) or whether it preferentially occurred in specific habitats.

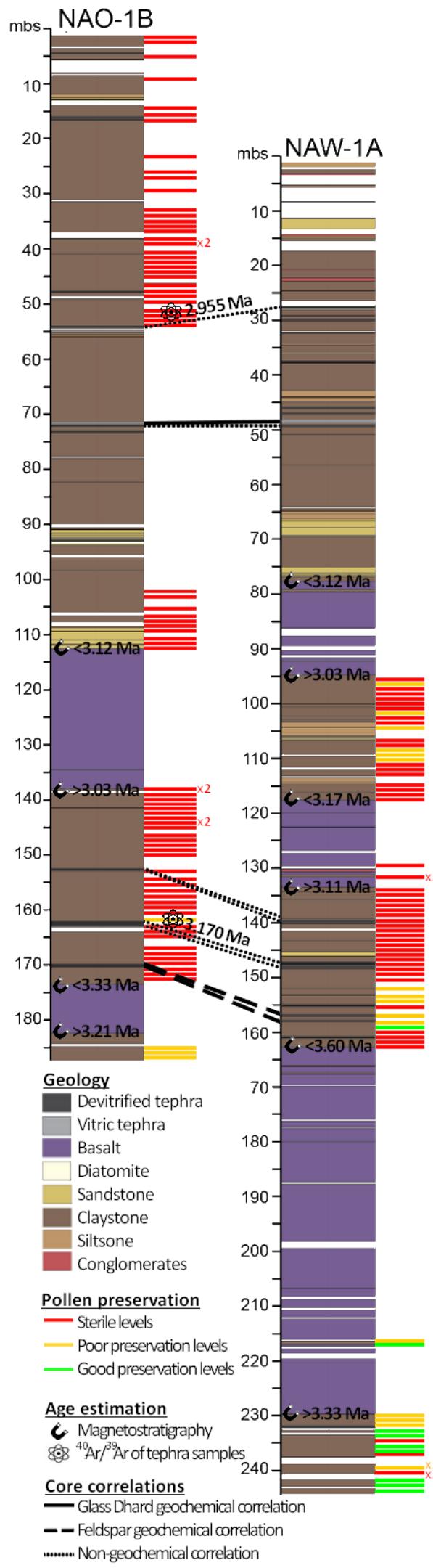
### 3.2. Material and Method

For this study, 175 pollen samples from two large cores drilled in the Ledi-Geraru research area (between the central and northern sub-area) of the Hominin Sites and Paleolakes Drilling Project (HSPDP) (Campisano et al., 2017) were analyzed for pollen content (**Figure 3.1**). These cores were drilled at the assumed position of the Hadar paleolake (Campisano, 2012). Of the 175 samples, 77 are from NAW14-1A core (ca. 245 m, 11°19'31.26" N, 40°45'53.68" E; NAW=Northern Awash Woranso) and 98 are from NAO14-1B core (ca. 187 m, 11°18'54.65" N, 40°44'12.80" E; NAO=Northern Awash Osolsi) (**Figure 3.3**) (Mohan et al., 2016). Preliminary  $^{40}\text{Ar}/^{39}\text{Ar}$  dating of the tephra in these two cores indicate an age of ca. 2.91 Ma for the tops of NAW14-1A and NAO14-1B cores (Garello, 2019). For the bases, paleomagnetism indicates an age older than 3.21 Ma

for NAO14-1B and older than 3.33 Ma for NAW14-1B (C. Campisano, personal communication).

The preparation of these HSPDP samples followed a standard pollen extraction protocol. 1) Sediment samples (5–30 g, depending on the sediment that was available) were treated with HCl (33%, 1 h) to remove carbonates; 2) digestion of silicate minerals in HF (48%, 12 h); 3) immersion in HCl (33%, 4 h) to remove fluorosilicates produced in the previous step; 4) immersion in KOH (20%, 10 min in a hot water bath) to remove organic matter; 5) sieving at 150 µm; 6) pollen stain with safranin and ethanol; (7) for productive samples, heavy liquid separation using sodium polytungstate set at a specific gravity of 2.2 grams per cm<sup>3</sup> to improve pollen concentration. For productive samples, the entire residue sample was analysed under a microscope. Slides were prepared with glycerin and sealed with Entellan. They were scanned at a magnification of ×250 and identified at a magnification of ×1,000. Pollen identifications (**Annexe table 3.1**) are based on the pollen reference collection available at *Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement* (CEREGE), Aix-en-Provence and published pollen atlases.

In addition to these data, 33 previously published pollen spectra from outcrops located near the drilling sites are also included. Four spectra dated “lower” Pleistocene come from the research area of Gona (Bonnefille et al., 1987; López-Sáez and Domínguez-Rodrigo, 2009), 26 spectra from the “upper” Pliocene of Hadar research area (Bonnefille et al., 1987, 2004), and two spectra come from the “lower” Pliocene and “upper” Pliocene sediments of Dikika research area (Bonnefille et al., 1987 ; Bonnefille, 2010). One spectrum comes from the “lower” Pliocene sediments of the MDR locality in the Woranso-Mille research area (Saylor et al., 2019) (**Figure 3.1** and **Table 3.1** ). The samples considered in this study are grouped into three geographical areas separated from each other by 25 km to 45 km : (1) the Woranso-Mille area is represented only by the MRD-210 sample, (2) the HSPDP NAW14-1A and NAO14-1B coring area represented by samples from this coring (all samples starting by “NAW14-1A” or “NAO14-1B”), and (3) the area comprising the Gona, Hadar, and Dikika research areas represented for the other samples (**Figure 3.1** and **Table 3.1**).



**Figure 3.3** - Position of the analyzed levels in the NAO-1B and NAW-1A cores with the degree of pollen preservation.

Stratigraphic positions	Labels	Research area	Sampling locality	Old age (Ma)	New Age (Ma)	Submembers based on old age	Pollen gains used	Pollen types used	Pollen-inferred paleovegetation	Confidence of pollen inferences
1	HAS-a1 <sup>1</sup>	Gona	HAS site	~2.35	~2.35	MMB assemblage	885 (99.2%)	31 (96.9%)	We-river	moderate
2	OGS-7(2) <sup>1</sup>	Gona	OGS-7 site	~2.60	~2.53 - 2.58±0.15	KH-7?	349 (100%)	20 (100%)	Bd-spring 1	high
3	OGS-7(1) <sup>1</sup>	Gona	OGS-7 site	~2.60	~2.53 - 2.58±0.15	KH-7?	364 (100%)	20 (100%)	Wcd-river 2	high
4	OGS-6a <sup>4</sup>	Gona	OGS-6 site	~2.60	~2.53 - 2.58±0.15	KH-7?	333 (100%)	20 (100%)	Wcd-river 2	high
5	5SN <sup>1,2</sup>	Hadar	Kada Hadar	2.95	2.96±0.01	KH-2	830 (99.9%)	26 (96.3%)	Wcd-river	high
6	NAW14-1A-158.90mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.19-3.20	§	274 (100%)	9 (100%)	Wcd-river 1	moderate
7	116f <sup>1,2</sup>	Hadar	Kada Hadar	<3.18	3.216 - 3.236	DD-2 or DD-1	398 (100%)	12 (100%)	Bd-river	high
8	116e <sup>1,2</sup>	Hadar	Kada Hadar	3.22	3.236 - 3.256±0.018	DD-2	399 (100%)	6 (100%)	Bd-river	high
9	333d <sup>1,2</sup>	Hadar	Kada Hadar	3.22	3.256±0.018	DD-2	443 (100%)	5 (100%)	Bd-river	high
10	333c <sup>1,2</sup>	Hadar	Kada Hadar	3.22	3.256±0.018 - >2.28	DD-2	665 (100%)	10 (100%)	Wcd-river 2	moderate
11	333b <sup>1,2</sup>	Hadar	Kada Hadar	3.22	3.256±0.018 - >2.28	DD-2	937 (100%)	5 (100%)	Bd-spring 1	high
12	134b <sup>1,2</sup>	Hadar	Kada Hadar	2.23	3.256±0.018 - >2.28	DD-1	2319 (100%)	32 (100%)	S-river	high
13	266b <sup>1,2</sup>	Hadar	Kada Hadar	3.24	3.256±0.018 - >2.28	DD-1	636 (100%)	3 (100%)	Bd-river	moderate
14	228b <sup>1,2</sup>	Hadar	Kada Hadar	3.25	>3.28	DD-1	359 (100%)	9 (100%)	Wcd-river 2	moderate
15	228a <sup>1,2</sup>	Hadar	Kada Hadar	3.25	3.3±0.04	DD-1	321 (100%)	10 (100%)	We-river	moderate
16	280Nj <sup>1,2</sup>	Hadar	Unda Hadar	<3.28	3.3±0.04 - 3.341±0.04	SH-3 or SH-4	382 (100%)	17 (100%)	Fb/Wd-river	moderate
17	KMm <sup>1,2</sup>	Harda	Hurda	3.37 - 3.35	3.341±0.04	SH-2	1813 (99.9%)	54 (98.2%)	We-river	moderate
18	KMl2 <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.343±0.03	SH-2	857 (100%)	33 (100%)	Wcd-river 1	high
19	KMl1 <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.343±0.03	SH-2	866 (99.7%)	27 (90.0%)	Fb/Wd-upper	moderate
20	KM <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.345±0.03	SH-2	491 (99.8%)	18 (94.7%)	Fb/Wd-river	moderate
21	KMj <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.350±0.03	SH-2	789 (100%)	22 (100%)	We-river	moderate
22	KMl <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.353±0.03	SH-2	1123 (100%)	16 (100%)	Bd-spring 1	high
23	KMh <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.356±0.02	SH-2	598 (100%)	17 (100%)	Bd-river	high
24	KMg <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.357±0.02	SH-2	503 (100%)	16 (100%)	Wcd-river 2	high
25	KMf <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.359±0.02	SH-2	499 (100%)	10 (100%)	Bd-spring 1	high
26	KM <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.360±0.02	SH-2	652 (100%)	12 (100%)	S-river	high
27	KMd2b <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.361±0.02	SH-2	690 (100%)	15 (100%)	Bd-spring 1	moderate
28	KMd2a <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.361±0.02	SH-2	1039 (100%)	10 (100%)	Bd-spring 1	high
29	KMc <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.370±0.01	SH-2	1232 (100%)	14 (100%)	We-river	moderate
30	KMb <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.372±0.01	SH-2	638 (100%)	7 (100%)	Bd-spring 2	moderate
31	KMa <sup>1,2</sup>	Hadar	Hurda	3.37 - 3.35	3.374±0.01	SH-2	1278 (100%)	9 (100%)	Bd-spring 1	high
32	NAW14-1A-217.02mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.39±0.03 - 3.41±0.03	§	154 (100%)	14 (100%)	S-xids	moderate
33	NAW14-1A-233.00mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.39±0.03 - 3.41±0.03	§	154 (100%)	14 (100%)	S-xids	moderate
34	NAW14-1A-233.84mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.39±0.03 - 3.41±0.03	§	187 (100%)	17 (100%)	S-xids	high
35	NAW14-1A-235.99mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.39±0.03 - 3.41±0.03	§	86 (100%)	26 (100%)	Bd-spring 1	high
36	NAW14-1A-236.82mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.39±0.03 - 3.41±0.03	§	135 (100%)	15 (100%)	Bd-river	high
37	339b <sup>1,2</sup>	Dikika	Qudaleita	3.4	3.42±0.03	UBM	163 (100%)	16 (100%)	S-xids	low
38	NAW14-1A-242.55mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.42±0.03	§	962 (99.9%)	50 (98.0%)	Wcd-river 2	low
39	NAW14-1A-243.76mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.42±0.03 - 3.43±0.03	§	257 (97.7%)	23 (92.0%)	Bd-river	high
40	NAW14-1A-244.08mbs <sup>this study</sup>	Ledi-Geraru	Core NAV14-1A	§	3.42±0.03 - 3.43±0.03	§	269 (97.1%)	21 (87.5%)	Bd-river	high
41	MRD-210 <sup>5</sup>	Woranso-Mille	MRD site	~3.8	~3.8	BBM	220 (100%)	29 (100%)	Bd-wd 2	high
42	Meshellu <sup>3</sup>	Dikika	Meshellu	~4	3.8-4	BBM	1519 (100%)	28 (100%)	Bd-spring 1	high

**Table 3.1** - Localisation, stratigraphy and results of V and V rivers scores. MMB assemblage: Maka'amitalu Basin assemblage ; KH-7: KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member. (1: Bonnefille et al., 1987, 2: Bonnefille et al., 2004, 3: Bonnefille, 2010, 4: López-Sáez and Domínguez-Rodrigo, 2009 and, 5: Saylor et al., 2019. For abbreviations of inferred paleovegetation, see Table 3.2.

To allow the comparison of all these pollen samples published at different times, we placed all pollen samples in the composite stratigraphic section of the Hadar and Busidima Formation (Campisano, 2007; Campisano and Feibel, 2008) The stratigraphic positions and ages of the 33 previously published pollen samples were reworked using the most recent stratigraphic data and dating (Campisano, 2007; Campisano and Feibel, 2008; Garello, 2019; Roman et al., 2008; Semaw et al., 2003; Wynn et al., 2006) (**Annexe table 3.2.a**). For the Hadar samples collected by the Bonnefille's team in the 1970s, we also used the archives stored at CEREGE, which provide a high degree of precision for their stratigraphic positions (**Annexe figure 3.3** and **Annexe table 3.2.b**). For the HSPDP data, the results of age estimates presented in this article were used (cf., **Part 3.3.2**) because not all tephra marker levels were identified in the cores (Garello, 2019).

We used the pollen score calculation approach of part 2 of this thesis, based on 249 modern pollen samples (soil and surface mud), to attribute local vegetation, regional vegetation and deduce the values of climatic parameters to the fossil pollen samples of this study. This method allows the calculation of two scores: a score for the identification of local and regional vegetation (V-score) and a score for the refined identification of local and azonal vegetation on dynamic deposition environments previously identified by the V-score (V-river score). In detail, the V-score allows the identification of vegetation types or sub-types (zonal and azonal) that originate from less dynamic deposition environments (e.g., soil, lake edges, spring water ponds) and summarily identifies those that originate from dynamic deposition environments (e.g., riverbeds or lake edges) (**Annexe figure 3.1** and **Annexe figure 3.2**). Samples identified as dynamic deposits by the V-score will then be reclassified as riparian vegetation types or sub-types using the V-river score. The V-score uses the pollen counts on relative abundance scale (based on the values in **Part 2 of this thesis**) and the river V-score uses the pollen counts on presence-absence. This approach allows the identification of 24 vegetation (including nine local riparian and azonal vegetation) divided into eight types of regional vegetation (**Table 3.2**). To use these approaches, a pre-processing of the data must be performed. Before the transformation of the pollen data into abundance scale and presence-absence, the undetermined pollen grains and the spores were removed, and the pollen types were homogenized to those of the **Part 1 of this thesis** (**Annexe table 3.1**). Some pollen types identified in the samples of our study have no equivalence in the 249 modern samples used to calibrate the approach of the **Part 1 of this thesis** (**Annexe table 3.1**). These

pollen types were therefore removed from our dataset only after the transformation of the pollen data to abundance scale and presence-absence. The number of grains and pollen taxa used in the calculation of scores is shown for each sample in **Table 3.1**.

Regional vegetation types identified by botanists <sup>0,2</sup>	Vegetation types (◦) and sub-types (•) that can be identified with the v and v-river scores approach <sup>10</sup>
Desert (D)	• Desert (D) <sup>0,2</sup>
Semi-desert grassland and shrubland (S)	<ul style="list-style-type: none"> <li>• Xerophytic dwarf shrubland<sup>3</sup> (S-xds)</li> <li>• Xerophytic shrubland<sup>3,6</sup> (S-xs)</li> <li>• Riparian very open woodland/shrubland along meandering perennial or seasonal river in S<sup>2,4,8</sup> (S-river)</li> </ul>
Somalia-Masai <i>Acacia</i> - <i>Commiphora</i> deciduous bushland and thicket (Bd)	<ul style="list-style-type: none"> <li>• Riparian dense <i>Tamarix</i> forest<sup>2,6</sup> (Bd-rtf)</li> <li>• Spring-fed woodland with <i>Acacia</i>, <i>Hyphaene</i> and <i>Tamarix</i><sup>2,4,5</sup> directly on springhead (Bd-spring 1)</li> <li>• Spring-fed woodland with <i>Acacia</i>, <i>Hyphaene</i> or <i>Phoenix</i><sup>4,5</sup> not directly on springhead (Bd-spring 2)</li> <li>• Steppe-like bushland with <i>Boscia</i><sup>2,4</sup> (Bd-wb 1)</li> <li>• Open woodland/bushland with Amaranthaceae and <i>Ocimum</i><sup>2,4</sup> (Bd-wb 2)</li> <li>• <i>Acacia</i> woodland and bushland<sup>2</sup> above 550m<sup>4,6</sup> (Bd-wb 3)</li> <li>• Moist edaphic <i>Acacia</i> woodland with abundant grasses and Cyperaceae<sup>2,3,4,5</sup> (Bd-wgc)</li> <li>• Riparian open <i>Acacia-Tamarix</i> forest surrounded by swamp along meandering river in Bd and upper limit of S<sup>2,4,8</sup> (Bd-river)</li> </ul>
Upland <i>Acacia</i> wooded grassland (We)	<ul style="list-style-type: none"> <li>• Riparian <i>Acacia</i> forest along straight river in We and upper limit of Bd<sup>2,4,8</sup> (We-river)</li> </ul>
Complex of afromontane undifferentiated forest with wooded grasslands and	<ul style="list-style-type: none"> <li>• Fb/Wd below 2500m without incursion of Ericaceous belt<sup>2</sup> (Fb/Wd-low/middle)</li> <li>• Fb/Wd above 2600m with incursion of Ericaceous belt<sup>2</sup> (Fb/Wd-upper)</li> <li>• Riparian afromontane undifferentiated forest along straight river in Fb/Wd<sup>2,4,8</sup> (Fb/Wd-river)</li> </ul>
Dry <i>Combretum</i> wooded grassland (Wcd)	<ul style="list-style-type: none"> <li>• Dry <i>Combretum</i> wooded grassland (Wcd)<sup>0,2</sup></li> <li>• Riparian open <i>Acacia</i> forest along straight river in Wcd<sup>2,8</sup> (Wcd-river 1)</li> <li>• Riparian close <i>Acacia-Cordia</i> forest along meandering river in lower limit of Wcd in contact with Bd<sup>2,8,9</sup> (Wcd-river 2)</li> </ul>
Afromontane rain forest (Fa)	<ul style="list-style-type: none"> <li>• Dry Fa<sup>2,7</sup> with PANN &lt; 1100mm/yr (Fa-sc 1)</li> <li>• Moist Fa<sup>2,7</sup> with PANN &gt; 1100mm/yr (Fa-sc 2)</li> <li>• Secondary moist evergreen afromontane forest<sup>2</sup> (Fa-w)</li> <li>• Upper transitional montane forest below E<sup>7</sup> (Fa-upper)</li> </ul>
Afroalpine vegetation and Ericaceous belt (A.E)	<ul style="list-style-type: none"> <li>• <i>Erica</i> bushland<sup>2</sup> or grove within <i>Hilechrysum</i> heath<sup>2,7</sup> (A.E-e)</li> </ul>

**Table 3.2** - Post-hoc vegetation with their abbreviation that can be identified with the V and V-river scores approach. 0: van Breugel et al. (2015), Micro-habitat types are based on 1: El-Moslimany (1983), 2: Friis et al. (2010), 3: Frey and Kürschner (1989), 4: *in situ* observations, 5: Barboni et al. (2019), 6: Bonnefille et al. (1987), 7: Mohammed et al. (2007), 8: Buchet (1982), 9: Carr (1998), and 10: Part 1 of this thesis.

Based on these scores and considering the significant intervals based on the 249 modern samples (see **Part 1 of this thesis**) (**Annexe figure 3.1** and **Annexe figure 3.2**), each fossil sample was assigned one of 24 possible vegetation (**Table 3.1**). Each sample was assigned the vegetation for which the sample V-score was highest among those in the range of significant values only. If there were none, then we considered the V-score between the range of significant and non-significant values. Finally, as a last resort, we considered the V-score in the range of non-significant values. If necessary, the same procedure was applied for V-river scores. We considered three levels of confidence for these attributions: 1) “high” for attributions based on a score only in the range of significant values; 2) “moderate” for attributions based on a score in the range between significant and non-significant values; 3) “low” for attributions based on a score only in the range of non-significant values. Vegetation types assigned with high confidence indicate that paleovegetation was not significantly different from the modern vegetation type assigned. Vegetation types assigned with moderate confidence indicates that the paleovegetation was different from the modern vegetation to a certain degree, but remain close to it. Vegetation types assigned with low confidence indicate that the paleovegetation was significantly different from the assigned modern vegetation, the paleovegetation then has no equivalent in the modern vegetation tested.

We used the definitions of Holdridge (1967) and Lugo et al. (1999) to describe climatic domains: tropical ( $TANN > 24^{\circ}C$ ), subtropical ( $TANN: 24-20^{\circ}C$ ), warm temperate ( $TANN: 20-12^{\circ}C$ ), super-arid ( $PANN: 0-125 \text{ mm/yr}$ ), per-arid ( $PANN: 125-250 \text{ mm/yr}$ ), arid ( $PANN: 250-500 \text{ mm/yr}$ ), semi-arid ( $PANN: 500-1000 \text{ mm/yr}$ ), and sub-humid ( $PANN: 1000-2000 \text{ mm/yr}$ ). Tree cover classes are defined as follows according to Edwards (1983) : open (tree cover: 1-10%), semi-open (tree cover: 10-25%), and semi-closed (tree cover: 25-50%), closed (>50%).

### 3.3. Results

#### 3.3.1. Stratigraphic position and age of the pollen samples

##### 3.3.1.1. Age adjustment for previously published samples

On the 33 pollen samples already published, we were able to adjust the ages of 31 samples. These new values are presented in **Table 3.1** and details of the references used for each age modification are in **Annexe table 3.2.a**. The adjustments may be corrections of interpretations, this is the case only for the 3 samples from the OGS-6 and 7 sites that López-Sáez and Domínguez-Rodrigo (2009) indicate as prior to Gauss-Matayuma polarity transition while Semaw et al. (2003) shows the opposite. There are also modifications related to the updating of stratigraphic knowledge and new dating which concerns the 15 samples from Kada Hadar, Unda Hadar, Oudaleita and Dikika. This is for example the case for sample 398b which was previously dated 3.40 Ma and is now dated 3.42 Ma based on the revised dating of Sidi Hakoma Tuff (Campisano, 2007). Note that for the Kada Hadar samples ranging from 116f to 266b, we have substituted the ages indicated by Bonnefille et al. (2004) by time intervals because the stratigraphic information published on these samples (Bonnefille et al., 1987; Bonnefille et al., 2004) does not allow such precision.

The age modifications for the 15 Hurda samples (the samples starting by "KM") was made from an age model presented in **Annexe table 3.2.b**. We took the liberty to propose an age model because the 15 Hurda samples were collected in the same stratigraphic section (Bonnefille et al., 1987) where sedimentation rate is considered regular (Taieb and Tiercelin, 1979). The age model is based on the following data: the unpublished stratigraphic section AV77 KM (=Hurda section) which gives the interval between the samples in this section (archives of palynological team of CEREGE : **Annexe figure 3.3**), the correlation of the thick coal/lignite layer of Hurda section that contains the sample KMe with the thin darker units within the KMT complex ( $3.36 \pm 0.02$  Ma) (Campisano, 2007), the new datation of Sidi Hakoma Tuff at  $3.42 \pm 0.03$  Ma (Campisano, 2007), and the fact that sample KMa was taken 26 m above the Sidi Hakoma Tuff (Bonnefille et al., 1987).

### 3.3.1.2. New Pollen Data from HSPDP

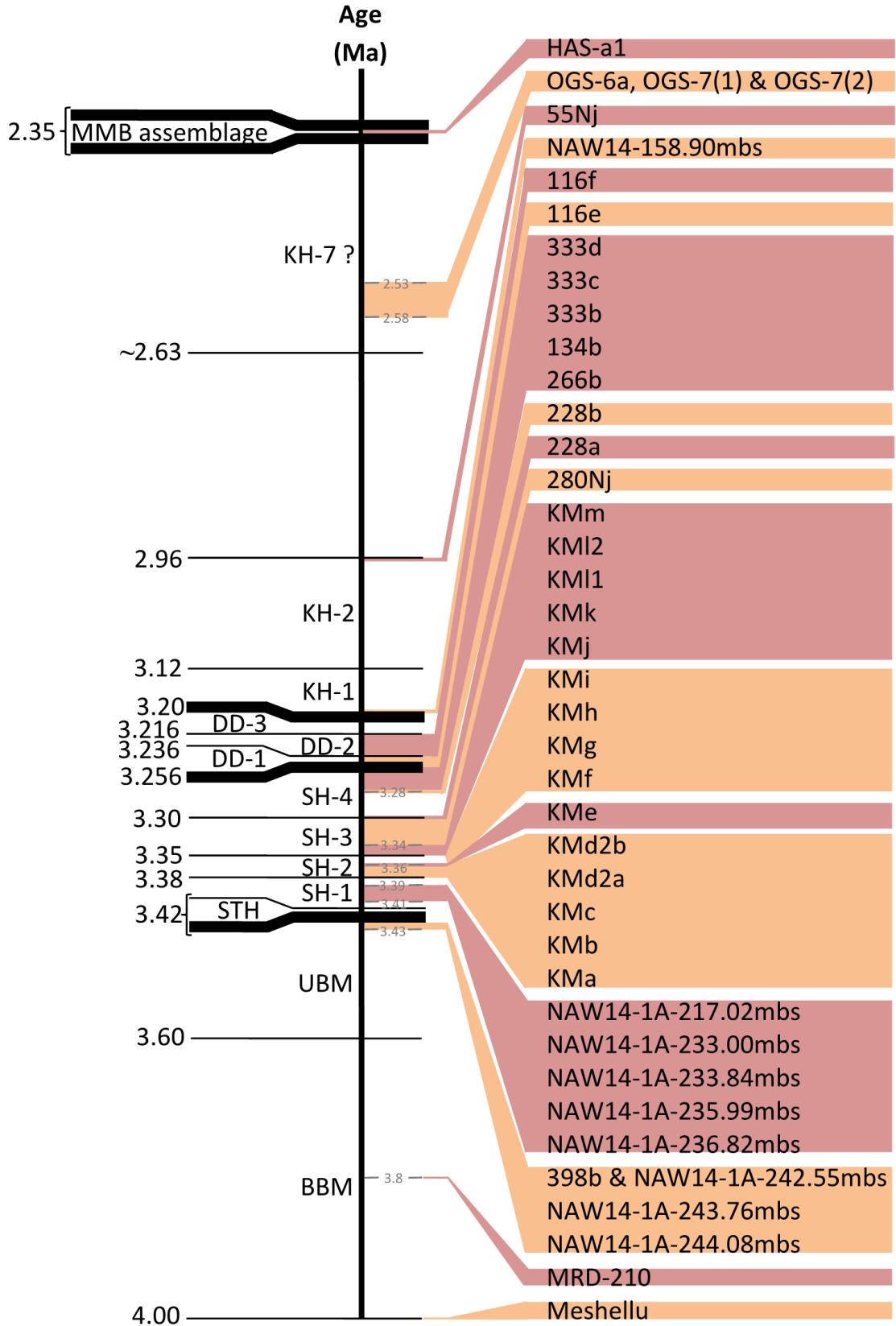
On the 175 samples of the NAW14-1A and NA014-1B core from the HSPDP, 144 samples (82%) were sterile or near sterile, 22 samples (13%) had very low pollen concentration with bad preservation and were not counted, and 9 samples (5%) had good pollen concentration with moderate to bad preservation. These 9 samples are the only ones that can be exploited. Using data of NAW14-1A and NA014-1B sediments, we investigated whether there was any correlation between the concentration and preservation of pollen grains and the characteristics of core sediments. But no correlation was found with any of the ratios and concentrations of chemical elements we tested (%Al, %Si, %K, %Ca, %Ti, %V, %Cr, %Mn, %Fe, %Ni, %Cu, %Zn, %Sr, Si/K, Si/Ti, Ca/K, Ca/Ti, K/Ti, Fe/Ti, Mn/Fe, Sr/Ca, Zr/Ti, Si/Ca, Ca/Ti, Sr/Ti, CaCO<sub>3</sub>), nor with isotopic ratios ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ), lithology, magnetic susceptibility, MSCL color reflectance, or loss on ignition (LOI at 550 °C and 1000 °C) (not shown).

The 9 exploitable pollen samples are from the base of the NAW14-1A core (**Figure 3.3**). Unfortunately, absolute dating of NAW14-1A is limited to two dates that do not allow building up of a robust age model. Using the age of  $2.955 \pm 0.037$  Ma obtained at 28.55 meters below surface (mbs) and  $3.169 \pm 0.009$  Ma at 147.7 mbs for the NAW14-1A core basaltes (C. Campisano, personnal communication), we estimated a mean sedimentation rate of 40.4 m/0.1 Ma. We used the presence of *Alangium* pollen type in sample NAW14-1A at 242.6 mbs (**Annexe table 3.1**) to correlate it stratigraphically with outcrop sample 398b (Bonnefille et al., 1987) whose stratigraphic position in the Hadar formation is well referenced (20 cm below tuff SHT, in UMB,  $3.42 \pm 0.03$  Ma) (**Annexe table 3.2.a**). This correlation is possible because *Alangium* is a pollen type that occurs exclusively in sample 398b. Using this information, the mean sedimentation rate and the age limitations given by the magnetostratigraphy of basalts (C. Campisano, personal communication) (**Figure 3.3**), we calculated time intervals for the 9 exploitable samples (**Annexe table 3.2.c**). We are aware that our estimates are approximate, especially for samples below 217 mbs. Further absolute dating is considered since 2015 at the *Géosciences Rennes* Laboratory of the University of Rennes 1 (France) by the HSPDP (C. Campisano, personal communication) but the timing of these dates is not yet decided. It will be necessary to compare our estimates with future absolute dates if they are realized.

### 3.3.1.3. Sample distribution in the Hadar and Busidima Formation

Knowing now the complete list of samples in this study and with our work of the sample ages, we now have an overview of our dataset in the composite stratigraphy of the Hadar and Busidima Formation (**Figure 3.4**). The samples are spread over several sub-members and can be subdivided into four large time sets : (1) between *ca.* 2.35 Ma and *ca.* 2.58 Ma for four samples contained in the samples contained in the Maka'amitalu Basin (MMB) assemblage and the sub-member Kada Hadar (KH)-7, (2) at  $2.96 \pm 0.01$  Ma in the top of sub-member KH-2 with a single sample, (3) between 3.216 Ma and  $3.43 \pm 0.03$  Ma in the sub-members from the base of the sub-member KH-1 to the top of the Upper Basal Member (UBM) with a total of 35 samples, and (4) between *ca.* 4 Ma and *ca.* 3.8 Ma in the base of Basal Member (BBM) with two samples. The new NAW14-1A samples provide the first pollen data on Sidi Hakoma (SH)-1 and add to the very sparse data from the top of the UBM.

Thirty-two of the 42 samples in our dataset document a 174 ka-long period from the top of UBM member to the top of SH-4 sub-member. In detail, the SH-4 sub-member is documented in its upper part by 6 samples covering a period of 24 000 years (or 24 ka). The SH-3 sub-member is well documented in its lower part by 4 samples covering a period of 7 ka, but its upper part has only 3 samples covering a period of 41 ka. The SH-2 sub-member is the best represented with 10 samples for 21 ka with intervals of 1 to 3 ka between each sample. Pollen samples from the SH-1 sub-member also offer a very good resolution with 5 samples for 2 ka-long period. Unfortunately, these data do not cover the base and top of SH-1. The Sidi Hakoma Tuff (SHT) is not documented, but the top of UBM is documented by 4 samples within 1 ka. The 9 remaining fossil pollen samples are sparse and loosely distributed in sub-members KH-7, KH-2, KH-1, DD-2 and BBM with 1 to 3 samples per sub-member, which limits the paleobotanical study of these sub-members. Finally, one pollen sample is available at the level of the MMB assembly. Despite the presence of just one sample, this punctual event is nevertheless relatively well-represented.



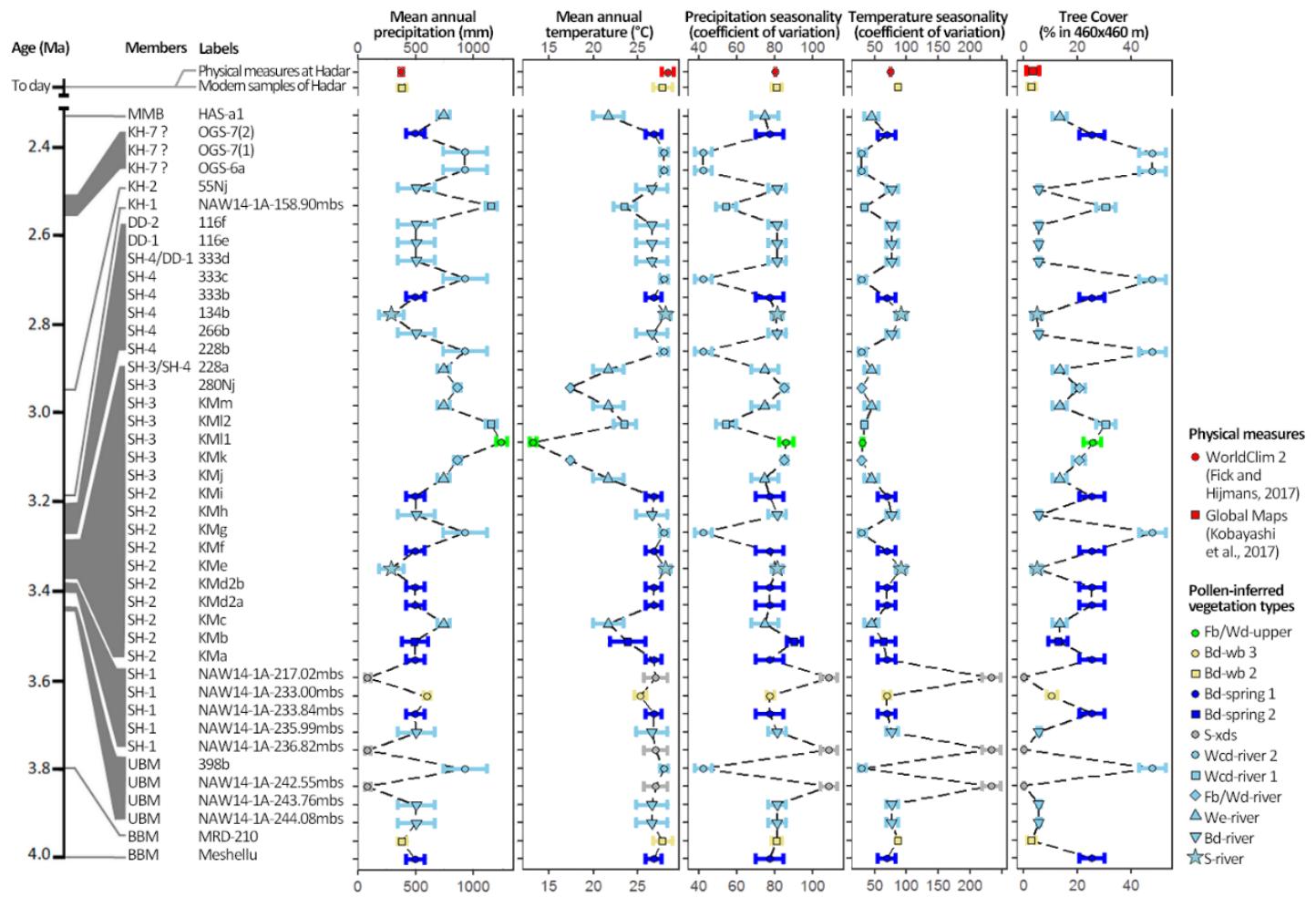
**Figure 3.4** - Position of the pollen samples in the composite statigraphy of the Hadar and Busidima Formations with respect to time scales. MMB assemblage: Maka'amitalu Basin assemblage ; KH-7: KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-3: Denen Dora Member sub-member 3 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; SHT: Sidi Hakoma Tuff ; UBM: Upper Basal Member ; BBM: Base of Basal Member.

### 3.3.2. Reconstruction of paleovegetation and paleoclimate

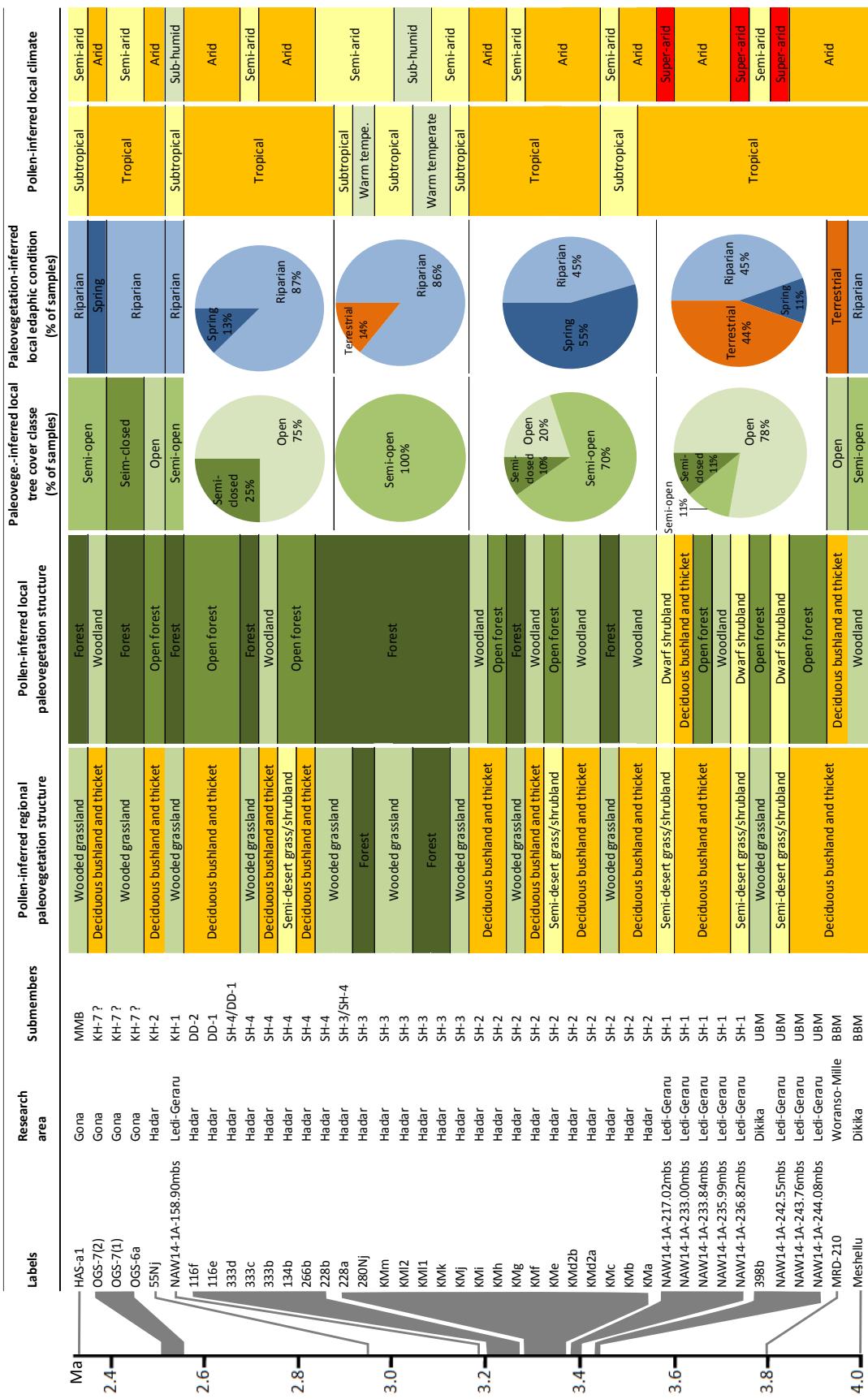
The paleovegetation reconstruction results and the climatic values obtained for each sample from the pollen data are given in **Figure 3.5**. A synthetic version of the results is given in **Figure 3.6**.

Using the method of the **Part 1 of this thesis**, we were able to propose regional and local vegetation for each of the 42 pollen samples (**Table 3.1**). These score calculations included 100% of the pollen grains and pollen types for 33 out of 42 samples (79% of the dataset). For each of the other 9 samples (21%), the scores could be calculated considering 82.0-99.9% of the pollen grains and 88-98% of the pollen types. Confidence in these attributions is high for 24 samples (57%), moderate for 15 samples (36%) and low for 3 samples (7%). The degree of confidence of the attributions is not randomly distributed but shows a relationship with the ages (**Table 3.1**).

Between ca. 4 Ma to ca. 3.8 Ma (BBM member), Dikika and Woranso-Mille punctual records indicate *Acacia-Commiphora* deciduous bushland and thicket type (Bd) linked to the Spring-fed woodland (Bd-spring 1 or 2) and Open woodland/bushland with Amaranthaceae and *Ocimum* (Bd-wb 2). Inferred climate is tropical arid. Inferred paleovegetation and climate are very similar to what is currently observed in this part of the lower Awash Valley (Friis et al., 2010a). From  $3.42 \pm 0.03$  Ma to  $3.39 \pm 0.03$  Ma (top of UBM et sub-member SH-1), at Dikika and at level of the site of core NAW14-1A, inferred paleovegetation remains of Bd type (except for sample 398b) linked to the Bd-spring 1, Riparian open *Acacia-Tamarix* forest surrounded by swamp along meandering river (Bd-river), and *Acacia* woodland and bushland above 550m (Bd-wb 3). The inferred paleoclimate is predominantly tropical arid with super-arid episodes which trigger the occurrence of semi-desert grasslands and shrublands (S) and xerophytic dwarf shrublands (S-xds) at the regional scale. In this interval, half of the samples indicate mostly “terrestrial” vegetation (44% of the samples), the other half indicate riparian vegetation (45% of samples). The inferred local paleo-tree cover indicated open vegetation for most of the samples (78% of samples).



**Figure 3.5** – Pollen-inferred climatic parameters and percent of tree cover (confidence interval of 95% ) based on supplementary data of the Part 1 of this thesis (Annexe table 2.3). MMB assemblage: Maka'amitalu Basin assemblage ; KH-7:KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member. For abbreviations of inferred paleovegetation, see Table 3.2.



**Figure 3.6** - Bilan of the evolution of climatic and environmental conditions. KH-7:KadaHadar Member sub-member 7 ;KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member.

From  $3.374 \pm 0.01$  Ma to  $3.353 \pm 0.03$  Ma in the sub-member SH-2 at Hadar, the pollen data reflect a paleovegetation dominated by Bd types. There is still an occurrence of regional vegetation of type S but linked to the Riparian very open woodland/shrubland along meandering perennial or seasonal river sub-type (S-river) which are linked to arid tropical episodes. Tropical super-arid episodes are replaced by semi-arid episodes. This is visible by the reconstruction of paleovegetation of Upland *Acacia* wooded grassland type (We) or Dry Combretum wooded grassland (Wcd) type linked to the Riparian close *Acacia-Cordia* forest along meandering river sub-type (Wcd-river 2). The paleoclimate remains tropical except at the base of sub-member SH-2 ( $3.372 \pm 0.01$  Ma to  $3.370 \pm 0.01$  Ma) where it becomes subtropical. The regional paleovegetation is mostly deciduous bushland and thicket and the local paleovegetation is dominated by woodlands that alternate with forests/open forests. The local paleo-tree cover is principally semi-open (70% of samples). The local vegetation signal alternates between riparian (45%) and spring-fed (55%).

From  $3.350 \pm 0.03$  Ma to  $3.3 \pm 0.04$  Ma in the sub-member SH-3 at Hadar, there is a major change. The pollen signal of the paleovegetation no longer shows Bd type, but an alternation of We and Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket type (Fb/Wd) with a bit of Wcd associated with the Riparian open *Acacia* forest along straight river sub-type (Wcd-river 1). This implies at the regional level an alternation of wooded grassland in subtropical climate with forest in warm temperate climate and semi-arid to sub-humid conditions. From the point of view of plant dynamics, it is a period marked by several episodes of expansion (compared to the present) of the mountain forest up to the foot horst and then of retreat of the mountain forest. During the retreat periods of montane forests, wooded grasslands mostly of the We type, which occur today in the central rift zone (280 km to the south-southwest of our sites). SH-3 sub-member exhibits the coldest and wettest conditions. Inferred paleoclimate is warm temperate and sub-humid at  $3.343 \pm 0.03$  Ma (KML1sample). At local level, the vegetation is mainly riparian forest (86%). For all samples, inferred paleo-tree cover is semi-open.

From ca. 3.28 Ma to ca. 3.216 Ma, in the upper half of the SH-4 sub-member to DD-2 sub-member at Hadar, the paleovegetation becomes mainly Bd type again with alternating Wcd related to the Wcd-river 2. The climate is tropical with a regional

paleovegetation alternating between wooded grassland in semi-arid condition to deciduous bushland and thicket in arid condition. The local paleovegetation is riparian (87%) with mainly open forest some occurrences of forest. The local paleo-tree cover is open (75%) or semi-open (25%). At ca. 3.19 M, in the base of the sub-member KH-1 in the NAW14-1A core, sample NAW14-1A-158.90 mbs reconstructs regional vegetation of Wcd associated with the Wcd-river 1 which indicates wooded grassland vegetation at regional level with riparian forest semi-open at local level. It indicates that the climate is once again subtropical subhumid for the last time in this recording.

At  $2.69 \pm 0.01$  Ma, the temporally isolated sample 55Nj from the top of the sub-member KH-1 at Hadar shows a Bd-river paleovegetation with climatic conditions tropical arid. The local paleo-vegetation is an open riparian forest. From  $2.58 \pm 0.15$  Ma to ca. 2.53 Ma in the sub-member KH-7 of Gona, 2 of the 3 samples indicate a paleovegetation of sub-type Wcd-river 2. The regional paleovegetation is therefore a wooded grassland and the local vegetation is semi-closed riparian forest in tropical semi-arid condition. The third sample (GSO-7(2)), the most recent, is still in tropical but arid condition with a regional paleovegetation of deciduous bushland and thicket and a local paleovegetation of semi-open woodland fed by spring. Finally, at  $2.69 \pm 0.01$  Ma in the MMB assembly of Gona, we have a sub-type of the We. The regional paleovegetation is a wooded grassland and the local vegetation is a semi-closed riparian forest in sub-tropical and semi-arid condition.

### 3.4. Discussion

#### 3.4.1. Importance of reinterpreting old pollen samples

The Meshellu pollen sample (ca. 4 Ma) was interpreted by Bonnefille (2010) like a swamp in "woodland-grassland landscape". This type of interpretation synthesizes the type of problem that we will find in this part of the discussion. Descriptions are too broad, too vague, and with paleovegetation structures that cannot be associated with existing vegetation. (Friis et al., 2010; *in situ* observation). This type of description characterizes the difficulty of separating the local from the regional pollen signal (Bonnefille et al., 1987). Here we have reinterpreted the sample as a local paleovegetation of semi-open

riparian woodland in a regional paleovegetation of deciduous bushland and thicket. Our reinterpretation remains in agreement with the great types of vegetation evoked by Bonnefille (2010) but the way of associating them changes. This is one of the interests of the score calculation approach of the **Part 1 of this thesis**. For the pollen samples 398b and MRD-210, our reinterpretations are in agreement with previous interpretations : deciduous bushland and thicket for MRD-210 (Saylor et al., 2019) and forest for 398b (Bonnefille et al., 1987; Bonnefille et al., 2004).

Pollen samples of the SH-2 sub-member (KMa to KMj) have previously interpreted as "delta swamp close to evergreen bushland and forest" (Bonnefille et al., 1987) or "close to wet/dry grassland" (Bonnefille et al., 2004). This is indicative of the second problem that encouraged us to use the scores of the **Part 1 of this thesis**, the subjectivity of interpretations. It can induce variations in environmental reconstruction from one article to another from the same authors. We agree with the idea of a delta in SH-2 which is also supported by geological observations (Wynn et al., 2008). In addition, our results propose an alternation of retraction and expansion phases of the active part of this delta based on the signal of the local vegetation which alternates between riparian vegetation (45%) and vegetation fed by springs (55%). Indeed, this indicates phases of retraction of the delta during which the springs are immersed allowing the development of spring vegetation and phases of expansion during which the springs are emerged. This hypothesis is based on the geological structure of the Awash Basin that favours the development of springs along the Awash River (Ayenew et al., 2008) and the presence of extensive tufas at Woranso-Mille and Gona at respectively 4.51-4.32 Ma et 3.5-3.7 Ma (Haile-Selassie et al., 2007; Semaw et al., 2005). For the local paleovegetation, we are also reconstructing riparian vegetation, but we do not agree with the term "swamp" which is too unclear and suggests open vegetation. Bonnefille et al. (1987) specifies that the SH-2 pollen percentages are similar to the pollen percentage recorded from Lake Turkana close to the delta of Omo river. This Omo area has a very heterogeneous tree cover between 20% and 80% (Global Maps, year 2008, 0.46×0.46 km :Kobayashi et al., 2017). This is consistent with our reinterpretations of local paleovegetation ranging between riparian open woodland to riparian semi-closed forest.

For pollen samples from the sub-member SH-3, previous reconstructions show a regional paleovegetation of forest (samples KMj à KMm), then woodland or wooded

grassland (samples 280Nj à 228a) (Bonnefille et al., 1987; Bonnefille et al., 2004). The authors of these reconstructions propose that the pollen signal of forest indicates an expansion of the montane forest to the base of horst (Bonnefille et al., 1987). We agree with this interpretation but only for the KMk, KMI1 and 280Nj samples based on our reconstruction of paleovegetation sub-types including into Fb/Wd type. But on the other levels, our study reconstructs a vegetation of wooded grassland type that marks the retreat of montane forest. The SH-3 sub-member therefore shows an alternation of montane forests and wooded grasslands rather than a simple development of montane forest with rapid replacement by wooded grassland on the end of the SH-3 sub-member (Bonnefille et al., 1987). At local level for the SH-3 sub-member, our pollen scores indicate that the local paleovegetation is semi-open riparian forest. There is no clear information in previous articles about local vegetation (Bonnefille et al., 1987; Bonnefille, 2010; Bonnefille et al., 2004).

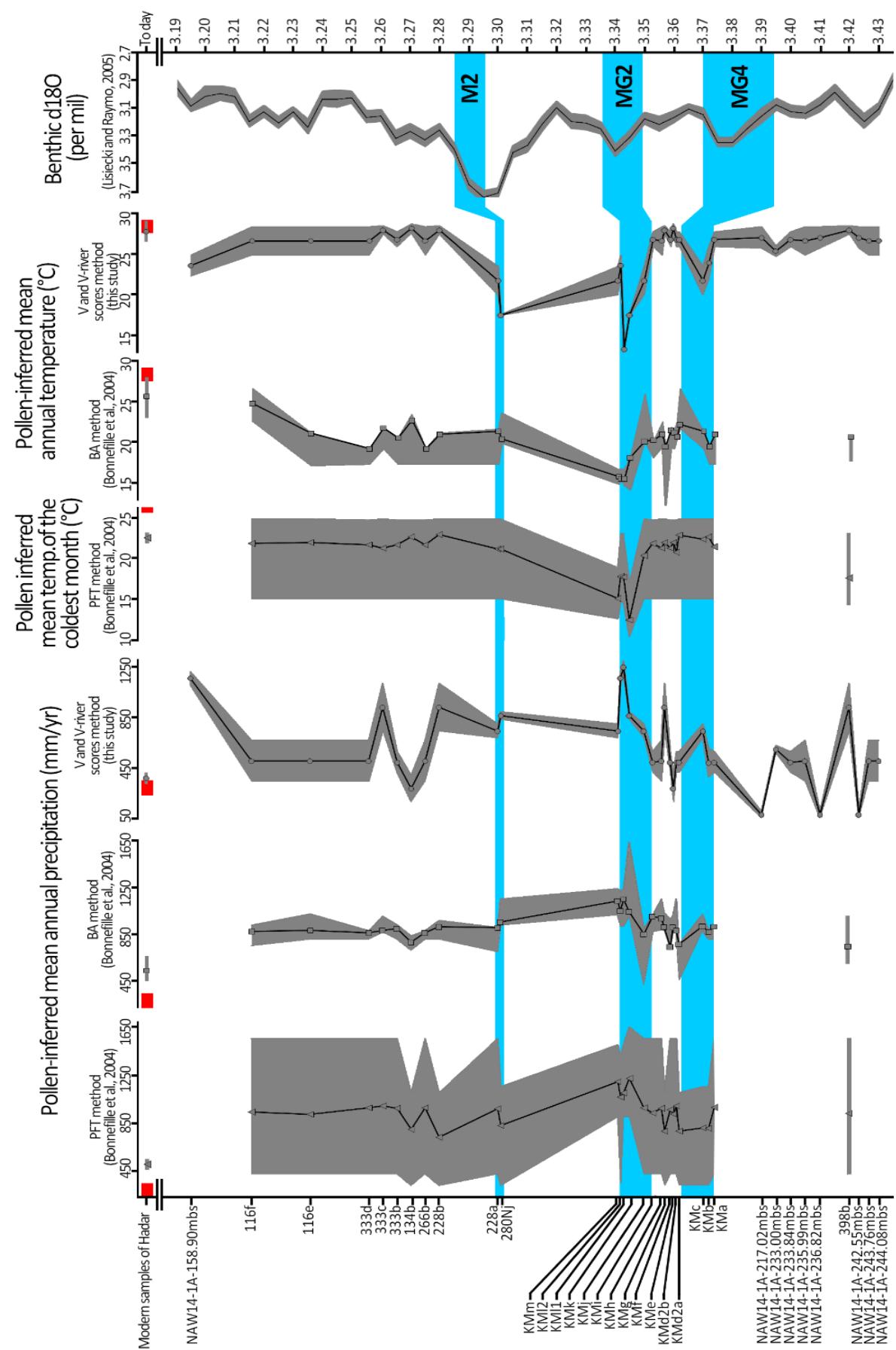
It is from the sub-member SH-4 that our reconstructions differ the most from previous ones. The earliest levels of the SH-4 sub-member (samples 228b and 266b) are seen regionally as a "grassland with sparse trees" (Bonnefille et al., 1987) or a woodland (Bonnefille et al., 2004). These reconstructions are still compatible with those in this article, which show a regional paleovegetation of wooded grasslands then deciduous bushland and thicket. But, on the next samples of the SH-4 sub-member (samples 134b to 333d), previous studies propose a regional vegetation of wet grassland (Bonnefille et al., 1987; Bonnefille et al., 2004). This is in contrast to our reconstitutions which give a regional paleovegetation dominated by deciduous bushland and thicket or semi-desert grass/shrubland. This vision of a regional wet grassland vegetation is based on the presence of lake sediment at the pollen sampling site and a high percentage of *Typha* pollen (Bonnefille et al., 1987) which are elements related to the local and not regional context. Regional reconstructions based on mammalian remains found in the SH-4 sub-member of Hadar propose a heterogeneous environment with "scrub-woodlands" and grasslands (Reed, 2008) which is in harmony with our results. For the local vegetation, there is still no clear interpretation to compare with ours. It is however important to note that pollens from *Hyphaene* and *Tamarix* showing a very strong local signal of edaphic woodland or forest (**Part 1 of this thesis**) have associated with "subdesert steppe" in sub-member SH-4 in previous studies (Bonnefille et al., 1987). This highlights the absence of scientific interest of previous studies to characterize the local riparian vegetation of Hadar.

which are almost always summarized at "swamp" (Bonnefille et al., 1987; Bonnefille, 2010; Bonnefille et al., 2004). This can be explained by the paradigms of their time, which privileged the research of drastic environmental changes to explain the evolution of the human lineage (Coppens, 1994).

For the sub-members DD-1 and DD-2, we reconstructed a regional paleovegetation of deciduous bushland and thicket against dry grassland vegetation in previous palynological studies (Bonnefille et al., 2004). At the level of Hadar faunal studies, a woodland for DD-1 and a wetland or floodplain for DD-2 are proposed (Reed, 2008). The faunal data therefore do not allow us to settle the question for DD-1 and DD-2. For KH-2, an evergreen bushland and forest (Bonnefille et al., 1987) or woodland (Bonnefille et al., 2004) had proposed at the regional level as we reconstruct a deciduous bushland and thicket at the regional scale and an open riparian forest at the local scale. Again, in view of the vegetation related to the different pollen types in previous pollen studies (Bonnefille et al., 1987), there was no attempt to separate the local riparian pollen signal from the regional signal. For the sub-member KH-7, we agree with the idea of a "wooded environment" at the regional scale (López-Sáez and Domínguez-Rodrigo, 2009) even if for us the most recent level still shows a more open and arid environment than the other. Finally, for the MBB assembly, a forest vegetation had been proposed at the regional level that we also find but at local level (Bonnefille et al., 1987). At the regional scale, we have rather a vegetation of wooded grassland.

### 3.4.2. Climatic reconstruction and evidence of MIS M2, MG2, and MG4

We compared climate reconstructions based on pollen data for three methods (**Figure 3.7**) : the Plant Functional Types (PFT) method (Bonnefille et al., 2004), the Best Analog (BA) method (Bonnefille et al., 2004) and the V and V-river scores method (**Part 1 of this thesis**) used in this article. In terms of range of precision, based on inferred values between 3.43 Ma and 3.19 Ma, the PFT method is the least accurate with a mean range of 962 mm/yr for the mean temperature of the coldest month and 9.4°C for the TANN. The BA and V and V-river scores methods have similar accuracies but are slightly better for the V and V-river scores method with respectively 225 mm/yr and 197 mm/yr for the PANN and 3.5°C and 2.3°C for the TANN. The real



**Figure 3.7** - Comparison of pollen-inferred climatic parameters for Plant Functional Types (PFT)(Bonnefille et al., 2004), la Best Analog (BA) (Bonnefille et al., 2004), and V and V-river scores methods (Part 1 of this thesis) with LR04 benthic  $\delta^{18}\text{O}$  stack (Lisiecki and Raymo, 2005). The red band represents the temperatures and precipitation measured at the Hadar site based on WordClim 2 (Fick and Hijmans, 2017).

today at Hadar. The BA method tends to overestimate the current PANN of 200 mm/yr and to underestimate the current TANN of 2.2°C. Using the V- and V-river score method, the values remain within the upper limit of the current range of measured PANN (400 mm/yr) and within the lower limit of the current range of measured TANN (27.5°C).

Except for a few points, there are the same trends in PANN and TANN estimates between the BA method and the V and V-river score method (**Figure 3.7**). But, climatic variations are more pronounced with the V and V-river scores method. By comparison of the V and V-river score climat data from this study with the LR04 benthic  $\delta^{18}\text{O}$  stack (Lisiecki and Raymo, 2005), it is possible to confirm a cooling linked to the Marine Isotope Stage (MIS) MG2 (ca. 3.34 Ma) at Hadar observed by Bonnefille et al.(2004) with TANN of ca. -15°C compared to the current one. It is also possible to highlight two other cooling episodes attributable to MIS M2 (ca. 3.29 Ma) and MIS MG4 (ca. 3.38 Ma) (considering the margin of error in dating samples) with TANN of respectively ca. -10.5°C and ca. -6.5°C compared to the current one. It is the paleo-temperatures inferred with the V- and V-river scores method that best follow benthic variations  $\delta^{18}\text{O}$ , still taking into account the error margins of dating, (Lisiecki and Raymo, 2005).

### 3.4.3. Variation of Intertropical Convergence Zone (ITCZ)

There are five recognized rainfall patterns in Ethiopia (Demissew et al., 2004; Diro et al., 2009, 2008), which constraint vegetation distribution (Friis et al., 2010a) (**Part 1 of this thesis**). Here we will use the nomenclature of Diro et al (2008). The majority (74%) of the vegetation reconstituted in this manuscript are sub-types of Bd, We, Fb/Wd and, S-river. Today these vegetation are related to the rainfall pattern IV that are currently being observed in the Northern Awash Basin and to the rainfall pattern III which is in central part of Ethiopia rift (**Part 1 of this thesis**). These two rainfall patterns are neighbors and are characterized by a similar seasonality with a main rainy season in July-August and a secondary rainy season in March-April. There are also 19% of the reconstituted vegetation is Wcd-river 1 and 2 which today can be seen in the south of Ethiopia in the area of rainfall pattern V (Diro et al., 2008). The rainfall pattern V differs from pattern III and IV in the distribution of rainy seasons which are in April-May and October (**Part 1 of this thesis**). The fact of finding a rainfall pattern V in the Northern Awash Basin implies a significant change in the annual rainfall distribution resulting in less seasonality (**Figure**

**3.5).** This implies a change in the seasonal oscillation of the Intertropical Convergence Zone (ITCZ) at  $3.42 \pm 0.03$  Ma,  $3.357 \pm 0.02$  Ma, ca. 3.28 Ma, ca. 3.19-3.20 Ma, and ca.  $2.53 - 2.58 \pm 0.15$  Ma. This type of variation may be related to the orbital eccentricity that influences the compression of the northern and southern limits of the ITCZ (Maslin and Trauth, 2009). Eccentricity also rhythms climate variability over cycles of  $10^4$ - $10^5$  ans (deMenocal, 2004; Trauth et al., 2007). This is consistent with our data, which can be related to periods of significant eccentricity. (Laskar et al., 2004). However, it is advisable to remain cautious on this last point in view of the margins of error in the dating. Finally, 7% of the reconstructions are of the S-xds sub-type, which is currently not represented in Ethiopia but is observed in Arabia at the northern limit of the ITCZ in summer. (**Part 1 of this thesis**). This would imply compression episodes of the ITCZ at least in its northern limit between  $3.39 \pm 0.03$  Ma and  $3.42 \pm 0.03$  Ma.

#### **3.4.4. The ecology of *Australopithecus afarensis* in the Northern Awash Basin.**

In the localities of Hadar and Dikika, *Australopithecus afarensis* is represented by 370 specimens distributed in all the sub-members of the Hadar Formation between sub-member SH-1 and sub-member KH-2 (3.42 Ma to 2.96 Ma). Outside of this interval, it was not observed (Campisano, 2007). It is also the only Hominine represented in the localities of Hadar and Dikika between 3.42 Ma to 2.96 Ma (Reed, 2008). The pollen data used in this study allow us to reconstruct paleoenvironments at Hadar and Dikika for the sub-members SH-2 to KH-2 (DD-3 and KH-1 excluded) and 25 km southwest of Hadar and Dikika for the equivalent sub-members SH-1 and KH-1 with the NAW14-1A core samples (**Figure 3.6**).

The paleovegetation reconstructed at Hadar and Dikika show significant climatic and floristic changes at the regional scale, in accordance with the palynological (Bonnefille et al., 1987; Bonnefille et al., 2004) and faunal (Reed, 2008) previously studies. This is what led these previous studies to characterise the species *Australopithecus afarensis* as eurytopic. (Behrensmeyer and Reed, 2013). However, the paleovegetation reconstructed here bring precisions at the local level from which a tendency emerges. At Hadar and Dikika, we observe the exclusive presence of woodlands or forests which are overwhelmingly riparian or fed by spring from sub member SH-2 to sub-member KH-2

(Figure 3.6). For sub-member SH-1, we do not have data directly from the Hadar and Dikika sites, but pollen records at Ledi-Geraru also show that 1/3 of the pollen samples are associated with woods or forests fed by river or spring. It is reasonable to think that these woodlands or forests linked to river or spring could also be present in Hadar and Dikika at the time of the sub-member SH-1. These woodlands or forests constitute stable micro-habitats at Hadar/Dikika despite the important paleoenvironmental and paleoclimatic variations at the regional scale thanks to the mitigate effect of the Awash paleoriver and the springs which partly fed it. (Friis et al., 2010a). This would explain the maintenance of *Australopithecus afarensis* at the Hadar and Dikika sites between 3.42 Ma and ca. 2.96 Ma despite the many climatic variations. In the Northern Awash Basin, there are two other sites with 3 occurrences of *Australopithecus afarensis* : Woranso-Mille at ca. 3.58 Ma (Haile-Selassie et al., 2010a) with a partial skeleton and Ledi-Geraru from the Denen Dora Member (ca. 3.256 Ma to ca. 3.200 Ma) with two molars (Wood, 2011). For Woranso-Mille, the reconstructed local paleovegetation is a riverine forest (Haile-Selassie et al., 2010b) but for the Denen Dora Member of Ledi-Geraru, there is no reconstruction of local vegetation. It is reasonable to propose that in the Northern Awash Basin, *Australopithecus afarensis* is a species that lives in tree-covered habitats of woodland or forest linked to river or/and spring.

As anticipated by Behrensmeyer and Reed (2013), it appears that it was indeed a problem of spatial resolution of the reconstructions that hindered the good understanding of the ecotonal/mosaic habitats of Hadar and Dikika. The association of *Australopithecus afarensis* with ecosystems of woodland or forest linked to river or spring is consistent because these micro-habitats with denser woody cover and higher plant diversity than the surrounding xerophytic vegetation are observed near rivers (Friis et al., 2010a). These riparian environments increase habitat diversity, spatial heterogeneity and species diversity (Reynolds et al., 2016). Moreover, although the existence of bipedalism for *Australopithecus afarensis* is clearly accepted, several morphological characteristics may suggest that its bipedal locomotion is not identical to bipedal locomotion of modern humans, mainly because of its marked arboreal character (Ward, 2002). However, it remains difficult to say whether these morphological characteristics linked to an arboreal locomotion are primitive features or real contemporary adaptations to *Australopithecus afarensis* (Haile-Selassie et al., 2010a; Ward, 2002). But, recent structural analyses of post-cranial bones clearly point in the direction of a contemporary

adaptation of *Australopithecus afarensis* for arboreal locomotion behaviour (Ruff et al., 2016). This is in accordance with our ecological reconstructions. Finally, the ecological studies of today's primates shows that few species are present in open or/and arid habitats. However, these species do not live directly in these open or/and arid environments but in the riparian forests that cross them (Reed and Bidner, 2004). This is, for example, the case of *Colobus guereza*, *Chlorocebus aethiops*, and, *Papio* spp. in Awash National Park (Ethiopia) (Dunbar and Dunbar, 1974). This also supports the hypothesis proposed here that *Australopithecus afarensis* has a riparian forest ecology in a regional environment that remains mostly dry and open between 3.42 Ma and ca. 2.96 Ma.

### 3.5. Conclusions

In the Northern Awash Basin, at level of *Australopithecus afarensis* sites that are commonly associated with ecotonal/mosaic habitats, our study shows that it is essential to correctly separate the local from the regional pollen signal. Indeed, riparian paleoenvironments allow the development of local and azonal paleovegetation very different from regional paleovegetation. The use of approach of the **Part 1 of this thesis** allows an objective and repeatable separation of the local and regional pollen signal. This has allowed a more accurate interpretations of the paleovegetation. This more accurate interpretations show for the first time an important diversity of azonal paleovegetation at Hadar and Dikika.

This new approach allowed us to propose for the first time a quantitative estimate of the local paleo-tree cover. It improved the estimation of climatic parameters from pollen, confirmed to Hadar and Dikika the impact of the global cooling phases associated with MIS MG2 and highlighted the impact of the global cooling phases associated with MIS M2 and MG4. We have observed, as in previous studies, that at the regional scale, there is nothing in the paleoenvironmental reconstructions to characterize the ecology of *Australopithecus afarensis*. But, reconstructions at local level, where the fossils of *Australopithecus afarensis* are found, show a very clear ecological preference for the woodlands and forests.

This study is a first step towards a better understanding of the ecology of *Australopithecus afarensis*. However, the interpretations of the ecology of *Australopithecus afarensis* made here by the pollens will have to be confirmed by the study of faunal remains and *Australopithecus afarensis* fossils. This is to ensure the solidity and consistency of the interpretations presented here.

### **3.6. Acknowledgments**

This work was supported by Aix-Marseille University (doctoral fellowship to BB), the Agence Nationale de la Recherche (HADoC Project to DB, ANR-17-CE31-0010-02). The authors extend gratitude to the entire HSPDP team and particularly the Northern Awash drilling and core description teams. We thank Andrew Cohen, Chad Yost, and Christopher Campisano from the University of Arizona for their support in this project and for providing the samples and data from the HSPDP cores of North Awash.

#### **4. Enregistrement pollinique et dynamique des microhabitats associés à la résurgence d'eau douce souterraine de Kisima Ngeda, lac Eyasi, Tanzanie**

• Titre de l'article : Vegetation dynamics of Kisima Ngeda freshwater spring reflects hydrological changes in northern Tanzania in direct connection with the Indian Ocean Dipole over the past 1200 years

• Auteurs : Benjamin Bourel<sup>1</sup>, Doris Barbóni<sup>1</sup>, Andrea M. Shilling<sup>2</sup>, Gail M. Ashley<sup>2</sup>

<sup>1</sup>CEREGE, Aix Marseille Université, CNRS, IRD, INRAE, Coll. France, Technopole Arbois, 13545 Aix en Provence cedex 4, France

<sup>2</sup>Department of Earth and Planetary Sciences, Rutgers University, Piscataway, 08854, NJ, USA

• Revue : Soumis à *Palaeogeography, Palaeoclimatology, Palaeoecology*

• Résumé : In the northern edge of saline Lake Eyasi, while vegetation is semi-desert, the modern spring of Kisima Ngeda (KN) sustains an *Acacia-Hyphaene* palm woodland and *Typha* swamps. To study the dynamics of this spring, which represents a plausible modern analog for the fossil springs documented in the nearby paleoanthropological and archaeological sites of Olduvai Gorge, we analyzed the pollen content of a short sediment core that documents vegetation changes since the last 1250 years (from 760 AD to 2011 AD). Our results show that (1) *Hyphaene* palms, which require meso-halophytic soil conditions were most abundant at the coring site before 1196 AD, when groundwater of KN spring was likely lower than at present, allowing intrusions of saline lake waters. (2) Since 1196 AD the palm woodland was replaced by a Mimosaceae woodland while the increased presence of *Typha* and Cyperaceae pollen also marks the increasing presence of wetlands in the area surrounding the site. (3) Since 1677 AD, the relative abundances of *Typha* and Cyperaceae indicate that groundwater level of KN spring is increasing and is now at its highest level in the last 1250 years. (4) The low abundance of *Typha* pollen mirrors paleoprecipitation changes reconstructed elsewhere in East Africa during the Medieval Climate Anomaly (MCA) and the Little Ice Age (LIA). We conclude that the Kisima Ngeda hydrological system, which has been active for more than 1250 years, responds rapidly to regional climate change, but is also capable of remaining active during

arid 250 years-long events such as the Little Ice Age. Our results support the hypothesis that this type of system helped to maintain hominin populations and activities in the arid lowlands of the rift on a multi-decennial scale.

- Mots clefs : Paleovegetation, Hydrosystem, Hominin, Olduvai, Ngorongoro, East Africa

## 4.1. Introduction

The eastern branch of the East African Rift System (EARS) is characterized by lowlands formed by a succession of arid grabens where evapotranspiration largely exceeds precipitations and many lakes are saline or alkaline (Frostick, 1997). In these arid environments, vegetation structure and composition greatly varies according to soil water availability (Caylor and Shugart, 2006). Micro-habitats with denser woody cover and higher plant diversity than the surrounding xerophytic vegetation are observed near rivers, shallow aquifers, and groundwater sources (springs) (Friis et al., 2010b; Greenway and Vesey-Fitzgerald, 1969). In an arid landscape, riparian environments increase habitat diversity, spatial heterogeneity and species diversity (Reynolds et al., 2016).

In the geological record, tufas (fossil springs) are often associated with abundant faunal remains including hominins and artefacts (Ashley et al., 2009), which suggests that springs and their associated micro-habitats may have offered valuable resources to early humans (see review in Barboni et al. 2019). Ashley et al. (2009) showed an active hominin exploitation of spring resources at Olduvai in the early Pleistocene (~1.79–1.74 million years ago, Ma) with the presence of specific stone tools at the paleo-spring level. Magill et al. (2013) showed that freshwater springs were important for *Homo* and *Paranthropus* hominin species subsistence at Olduvai particularly during periods of low precipitation between 2.0-1.8 Ma, suggesting that places of foci like springs may have enhanced competition among species during dry periods (Barboni, 2014). Springs and their associated palm woodlands may also have played an important role in the ecology and locomotion of early hominins (Barboni et al., 2019). Springs may have served as hydro-refugia during arid climate periods (Cuthbert and Ashley, 2014), and facilitated geographical dispersion of hominins (Cuthbert et al., 2017). The Plio-Pleistocene record of springs, however, is discontinuous and lacks sufficient temporal resolution in any given area to study their dynamics and their impact on the ecosystems. How stable are the

micro-habitats associated with the springs on the pluri-decennial to pluri-centennial time-scale? How does the groundwater discharge evolve on human (*ca.* 100 year) timescales? Is it continuous and stable, continuous and variable, or temporary?

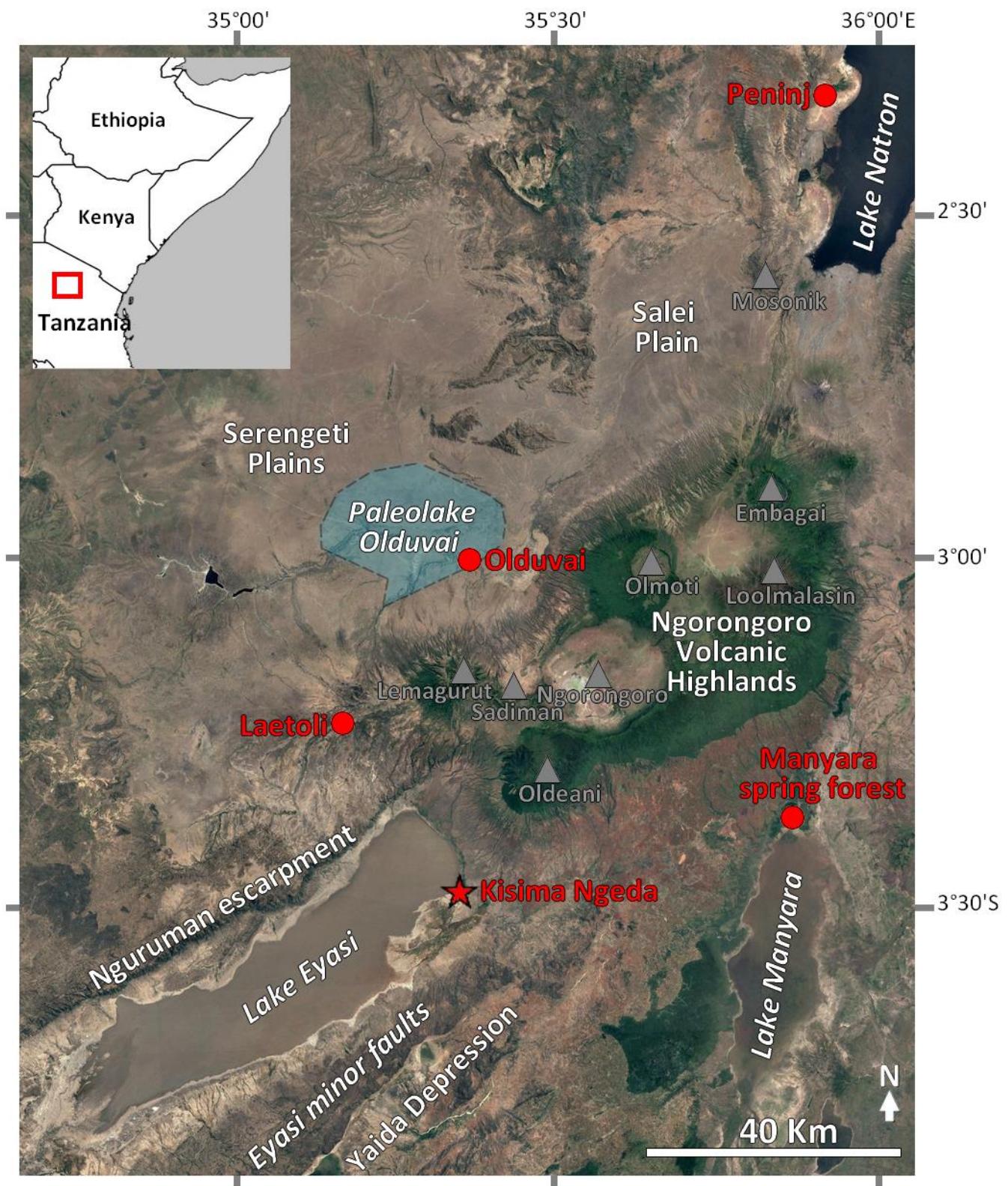
Modeling studies have partially answered these questions (Cuthbert et al., 2019, 2017). They showed that in the EARS groundwater discharge dynamics dependent on climate, topography, and the geological structure of the basement (Cuthbert et al., 2019), making reconstructions of groundwater discharge dynamics complex. Except for the eastern part of the Afar Triangle and the northern part of Lake Turkana, Cuthbert et al. (2019)'s model indicates that in the EARS the time required for a groundwater system to re-equilibrate itself after a climate change, which depends on the distance between recharge area and local aquifers, is about 10-1000 years, *ca.* 500 years on average. In fact, the grabens of the eastern branch of the EARS include finely intertwined zones of short and long groundwater response time of 10-100 yr or 100-1000 yr (Cuthbert et al., 2019). If a given spring responds to climate change with a delay of 100-1000 yr or more, than such spring may indeed have played a critical role in mitigating climate impact on hominin dispersion and ecology. On the contrary, if the delay is in the range 10-100 yr, then it is unlikely that the spring was responsible for the development or maintenance of hominin activities in lowlands of the rift during dry periods. In this later case, the co-occurrence of hominin remains and springs at a given archaeological site would have to be treated with caution as it may be potentially biased by taphonomic processes and/or fossil collection strategies (Behrensmeyer and Reed, 2013). Real-world data offering adequate temporal resolution are therefore needed.

Olduvai Gorge is a world-famous site located in the Crater Highlands at the southern edge of the Serengeti Plains (Leakey, 1971, 1966). The abundance of tufa mounts in paleosurfaces associated with anthropological and archaeological sites has been recognized at several Olduvai sites including FLK N (1.79 Ma) and FLK Zinj (1.89 Ma) (Ashley et al., 2010a, 2010b). Oxygen isotopic analyses of the tufas indicates groundwater was fresh (Ashley et al., 2010a), while paleolake Olduvai was saline/alkaline (Hay and Kyser, 2001). Phytolith and pollen data from FLK Zinj and FLK N paleosurfaces show that localized palm woodlands within an otherwise grass-dominated landscape were present near the springs (Arráiz et al., 2017; Ashley et al., 2010a; Barboni et al., 2010; Bonnefille, 1984). These paleo-springs have enabled the development of dense wooded vegetation

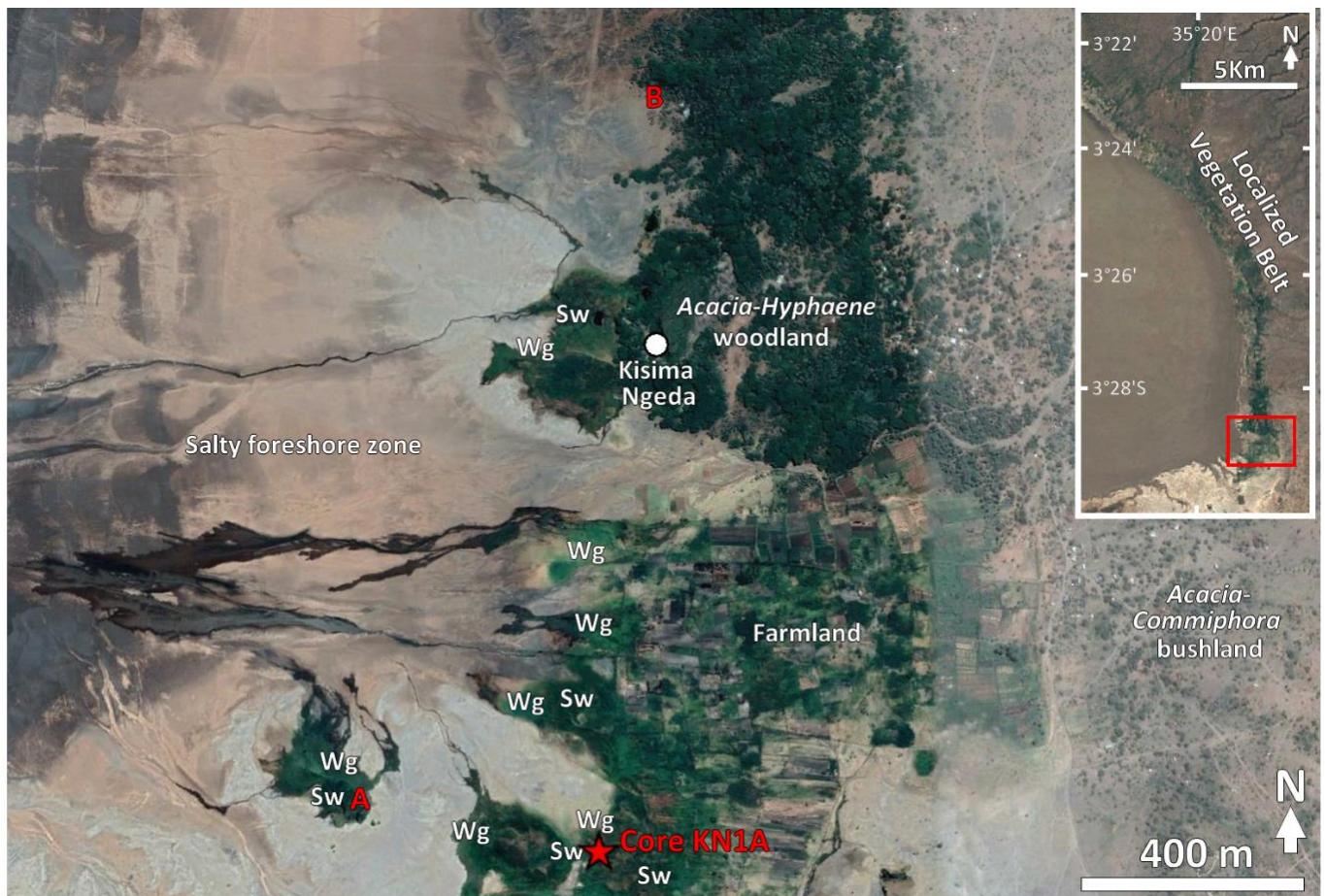
despite the semi-arid paleoclimate (paleoprecipitations of 250 to 700 mm/yr, Magill et al., 2013) and saline/alkaline waters of paleolake. The presence of paleo-springs is also suspected in the nearby paleontological sites of Laetoli and Peninj (Barboni 2014), and many springs are still active today in this region of northern Tanzania, such as at Esere near Laetoli (Barboni et al., 2019), and at the northern edges of Lake Manyara (Greenway and Vesey-Fitzgerald, 1969) and Lake Eyasi (Albert et al., 2015). Springs are therefore persistent hydrogeological features in the Crater Highlands region of north Tanzania, which modern representative may serve as analogs.

We focused our research on the spring of Kisima Ngueda locality at the northern edge of Lake Eyasi because this spring is associated to a palm woodland and is located at the edge of a saline/alkaline lake. Kisima Ngueda spring offers the opportunity to study the dynamics of an hydrosystem analog to the fossil springs of Olduvai Plio-Pleistocene hominins sites. Kisima Ngueda spring recharge occurs in the Ngorongoro Volcanic Highlands (Deocampo, 2002), like for Olduvai paleo-springs located about 50 km north of Lake Eyasi (Ashley et al., 2009). Rainwater infiltrates and is transported to the dry lowlands by moving groundwater. Groundwater flows down slope in the sub-surface within pyroclastic beds (tuffs) between relatively impermeable basaltic beds (Norton, 2019). Under pressure, groundwater discharges within alluvium at the surface or near the surface in the lowlands at the foothills of Ngorongoro Volcanic Highlands (Norton, 2019; Shilling, 2013), but it is not clear yet if groundwater discharges through faults, fractures, or dikes as suggested elsewhere in the Ngorongoro area (Deocampo, 2002), or where the aquifer intersects the surface due to the slope topography (Norton, 2019). Faults were not directly observed in the area of Kisima Ngueda (Pickering, 1964), but the continued hydrological activity may have altered fault contacts and fractures and made them hard to see on the surface (Deocampo, 2002). At the foothills of Oldeani volcano, water table is high, and a dense wooded vegetation belt with tall *Acacia xanthophloea* and *Hyphaene petersiana* palm trees develops despite the semi-arid climate (Fick and Hijmans, 2017) (**Figure 4.1** and **Figure 4.2**).

In order to study the influence of Kisima Ngueda spring on the vegetation, how this micro-habitat in the northeastern margin of Lake Eyasi possibly responded to past climatic changes, and its resilience, we analyzed the pollen content of a sediment core that was collected in one of the spring-fed swamps of this area (**Figure 4.1**).



**Figure 4.1** - Satellite image of sector around the Ngorongoro Volcanic Highlands. Grey triangles are volcanoes, red circles hominin paleontological sites, and red star is our study site. The approximate contour of paleolake Olduvai is for 1.75 Ma age (based on Barboni, 2014). Lake Emakat is the crater lake at Embagai.



**Figure 4.2** - Satellite images of the localized vegetation belt on the northeastern margin of Lake Eyasi with a zoom on the Kisima Ngeda area. Sw: swamp, Wg: wet grassland. The letters A and B are respectively the positions of the photo in Figure 3.A and the photo in Figure 3.B.

## 4.2. Study area

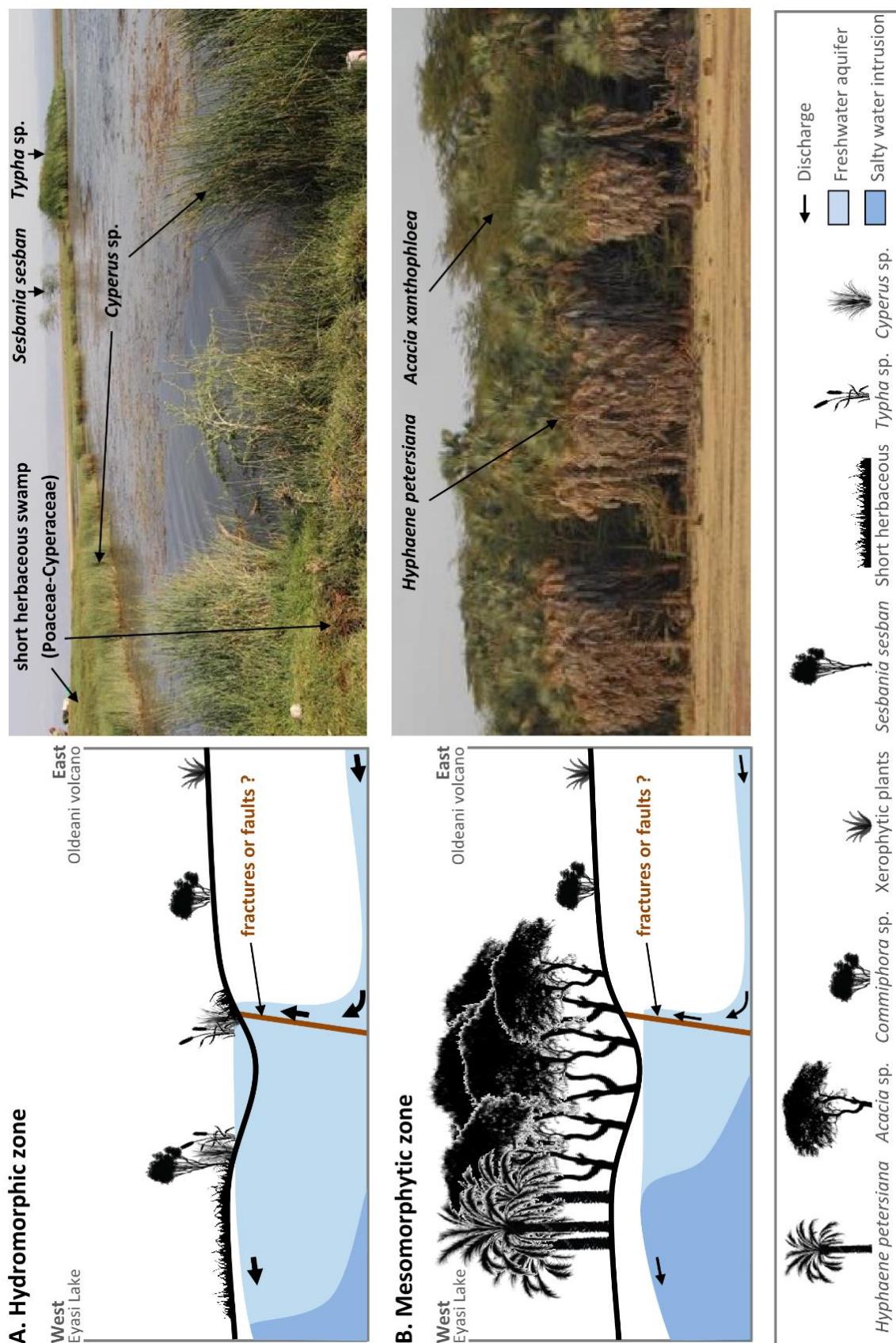
The eastern branch of the EARS, also called the Gregory Rift, contains a series of connected faults and grabens (Dawson, 2008). The Eyasi Basin located in the southern extremity of Gregory Rift between the Ngorongoro Volcanic Highlands and the Iramba Plateau is 160 km long and 30 km wide. It is composed of two sub-basins: The East Eyasi Basin with Lake Eyasi (100 km long) and the West Eyasi Basin with Lake Kitingiri (60 km long) (Ebinger et al., 1997). The East Eyasi Basin is very flat and lies at 1030 m above sea level (Foster et al., 1997). It is bordered by the Eyasi Fault and the Nuguruman escarpment to the west, the Ngorongoro Volcanic Highlands to the north, and minor

faulting of a basement warp to the east (**Figure 4.1**) (Foster et al., 1997(Foster et al., 1997). East Eyasi Basin has an half graben morphology; its modern position and morphology was likely acquired about 1 Ma ago (Foster et al., 1997). The northeast margin of the lake is bordered by the Neogene trachy-andesitic Oldeani volcano, which lies at 2800 m.

In the region encompassing East Eyasi Basin, Ngorongoro Volcanic Highlands, Salei Plains, and southern Serengeti Plains, the climate is very contracted due to significant relief variations. It is principally subtropical semi-arid in the lowlands, and warm temperate semi-arid in the highlands (Holdridge, 1967b)with mean annual temperatures (TANN) between 15°C and 24°C and mean annual precipitations (PANN) between 500 mm/yr and 900 mm/yr (Fick and Hijmans, 2017). At the level of the East Eyasi Basin, the climate is sub-tropical arid with TANN of 22-24°C and PANN of 550- 650 mm/yr. The Ngorongoro Volcanic Highlands allow the local development of a colder and wetter climate, with TANN between 12°C and 17°C and PANN between 1100 mm/yr and 1200 mm/yr. Moderate to low rainfall and intense evaporation of about 2500 mm/yr in the lowlands has allowed the development of the alkaline saline lakes in the region (Deocampo, 2005).

In the northeastern edge of Lake Eyasi, six vegetation zones can be defined based on the literature and our personal observations. (1) The Afromontane zone occurs above *ca.* 1600 m and 2450 m on the eastern and western wall of Mount Oldeani, respectively. It is constituted of Afromontane forest with trees of *Albizia gummifera*, several *Olea* species, and *Hagenia abyssinica* (Herlocker and Dirschl, 1972). The lower part of the Afromontane forest is delimited by an *Acacia lahai* high woodland belt. (2) The xerophytic zone covers most of the landscape below the Afromontane zone, and is constituted of *Acacia-Commiphora* bushlands. On the slopes of Mount Oldeani above *ca.* 1250 m, the *Acacia-Commiphora* bushlands include trees of *Commiphora madagascariensis*, *Acacia tortilis*, *A. mellifera*, and *A. senegal* and grasses are common (Herlocker and Dirschl, 1972). Below *ca.* 1250 m, bushes become more abundant than trees with decreasing elevation, and include *Acacia oerfota*, *Maerua trichophylla*, *Cordia rothii*, and *Commiphora* species. Grasses are scarce and plants more adapted to arid conditions are present such as *Adansonia digitata* (baobabs) and *Euphorbia candelabrum* (cactoid trees) (Herlocker and Dirschl, 1972). (3) The halo-xerophytic zone occurs on Lake Eyasi margins where salt

mudflats bare of vegetation are numerous. Vegetation on the lakeshore is scarce and exclusively constituted of salt loving species such as *Suaeda monoica*, *Volkensinia prostrata*, *Neuracanthus scaber*, and *Senecio mesogrammoides* (Rea, 1935). At the foothills of Mount Oldeani, between the *Acacia-Commiphora* bushlands and Lake Eyasi margin, a discontinuous vegetation belt can be identified on satellite images by its green color contrasting with the generally pale-yellow color of the surrounding xerophytic bushlands. This localized vegetation belt is formed by what we call a mesophytic zone, a halo-mesophytic zone, and an hydrophytic zone. (4) The mesophytic zone is a 0.5 km wide and ca. 7 km long discontinuous strip of groundwater-fed tall *Acacia xanthophloea* and *Hyphaene petersiana* palm woodland stretching along the north-eastern side of Lake Eyasi. In this part of the landscape, groundwater does not emerge but is close enough (ca. 1 m) to the surface to maintain dense woodlands with dense grass cover (*in situ* observation). (5) The halo-mesophytic zone is characterized by high concentration of *Hyphaene petersiana* palm trees (with more or less bare soil); *Hyphaene* favoring more salty/alkaline water than the *Acacia xanthophloea* trees, the halo-mesophytic zone occupies the edge of the spring woodland facing the lake margin (*in situ* observations) (**Figure 4.3**). (6) The hydrophytic zone corresponds to a complex of water ponds and swamps which develop where groundwater emerges at ground level. It is located between the halophytic and mesophytic zones. Cattail *Typha* sp. and sedges such as *Cyperus laevigatus* are present in the swamps along with the small water-loving tree *Sesbania sesban* (**Figure 4.2**). Short herbaceous swamps with sedges (probably *Cyperus laevigatus*) and grasses (e.g. *Sporobolus spicatus*, *Diplachne fusca*) and *Sesbania sesban* trees are also present (Hughes, 1992), as well as tussocks of the tall grass *Sporobolus consimilis* mixed with low *Cyperus* sp. sedges (probably *Cyperus laevigatus*) (**Figure 4.3**). The sediment core we studied was collected in this hydrophytic zone near to the locality of Kisima Ngeda.



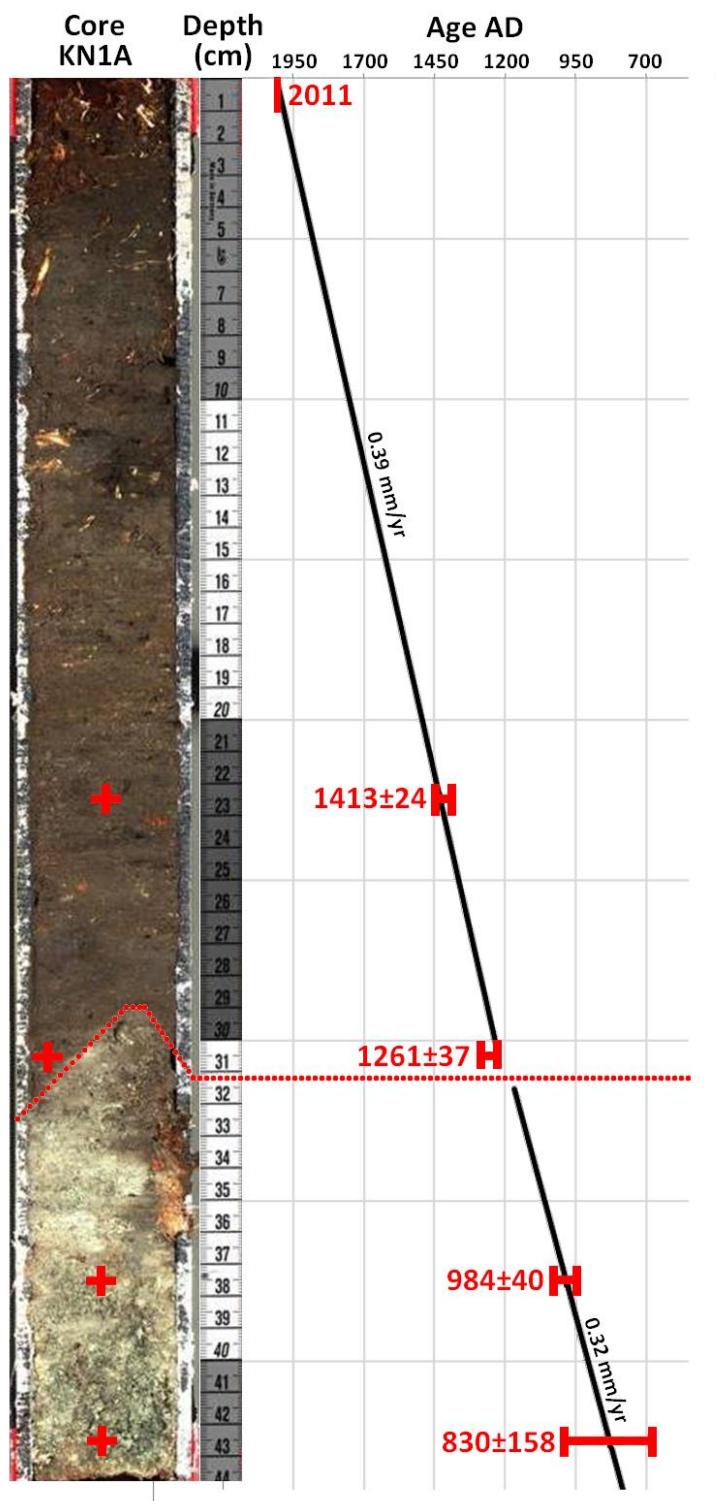
**Figure 4.3** - Theoretical drawing showing vegetation according to the groundwater level which controls the spatial extent of the hydrophytic (A) and mesophytic (B) zones. Photographs show how the two zones may look like in the area of Kisima Ngeda (photographs taken by D. Barboni and G.M. Ashley in 2011).

### 4.3. Material and Method

During the summer of 2011, we collected a 43 cm-long core in a *Typha* wetland near the groundwater-fed locality of Kisima Ngeda at 3°29'8.5"S and 35°20'53.5"E (**Figure 4.2**). Core KN1A was described as black peat with pieces of vegetation in the upper interval from 0 to 31 cm, and light grey silty clay without any visible pieces of vegetation in the lower interval from 31 cm to 43 cm (**Figure 4.4**). The contact of the upper and lower intervals is not horizontal, and may be the result of coring. A sub-sampling into sections of 1 cm was carried out (Shilling, 2013).

Pollen samples of 1 cm-thickness were taken every centimeter from the top to the base of the core. Level 30-31 cm was taken from the upper interval of the core. Pollen grains were concentrated by acid digestion using HCl (33%, 4h) to dissolve carbonates, HF (48%, 12h) and HCl to remove fluosilicates, and KOH (20%, 10') to remove humic acids. Sieving was done at 150 microns. Pollen grains were stained using safranin. Pollen identifications are based on the pollen reference collection available at CEREGE (Aix-en-Provence, France) and pollen atlases (e.g. (Bonnefille and Riollet, 1980). Pollen types are based on Vincens et al. (2007). Pollen percentages were calculated excluding undeterminable pollen grains and spores. In the lower part of the core, all samples have low (<110) pollen counts due to poor preservation **Annexe table 4.1**. These samples were not excluded from the analysis given their importance in documenting paleovegetation before the peat inception. However, we calculated confidence intervals associated with percentages of *Typha* in Figure 7 to consider the potential bias linked to such low pollen counts. Calculation of confidence intervals follows Suchéras-Marx et al. (2019). Pollen concentrations were not measured and results presented here are uniquely based on pollen taxa relative abundances. Pollen grains for 14C dating were obtained by the same chemical treatment, without using ethanol and safranin to avoid carbon contamination. The radiocarbon age of material was determined by Beta Analytic (Miami, Florida) and ARTEMIS instrument (Saclay, France) (**Annexe table 4.2**). Calibration of the radiocarbon ages was carried out with OxCal 4.4 program (Ramsey, 2001), using the IntCal13 calibration (Hogg et al., 2013). The IntCal20 calibration (Reimer et al., 2020) not yet available at the time of writing was not used. However, as shown in **Annexe table 4.2**, this new calibration brings little differences to our calibrated radiocarbon ages (**Annexe table 4.2**).

Pollen counts are given along with ecological and physiological information about the plant taxa they represent (**Annexe table 4.1**). Information about plant habit, plant preferred ecological zone, salt tolerance, life cycle (for herbaceous taxa), and leaf phenology (for arboreal taxa) were obtained from the literature (“African Plant Database,” 2019; Agnew and Agnew, 1994; Dale and Greenway, 1961; Greenway and Vesey-Fitzgeral, 1969; P. Loth and Prins, 1986) . The ecological and physiological traits indicated for the pollen taxa of KN1A core correspond to plant species potentially present in the study area. For example, *Albizia* pollen likely represent the deciduous *Albizia anthelmintica*, that is currently present in the spring woodland at the northern edge of Lake Manyara (Greenway and Vesey-Fitzgeral, 1969). Pollen percentages by ecological zones (defined in the study area part) and by life cycle/leaf phenology were calculated excluding pollen types with undifferentiated ecological zones, life cycle or leaf phenology (**Annexe table 4.1**). Confidence intervals associated with percentages of *Typha* in **Figure 4.7** were calculated following Suchéras-Marx et al. (2019).



**Figure 4.4** - Photograph of core KN1A cut in half with age models based on calibrated radiocarbon ages of 14C dating of concentrated pollen grains. Red crosses indicate position of samples for 14C dating.

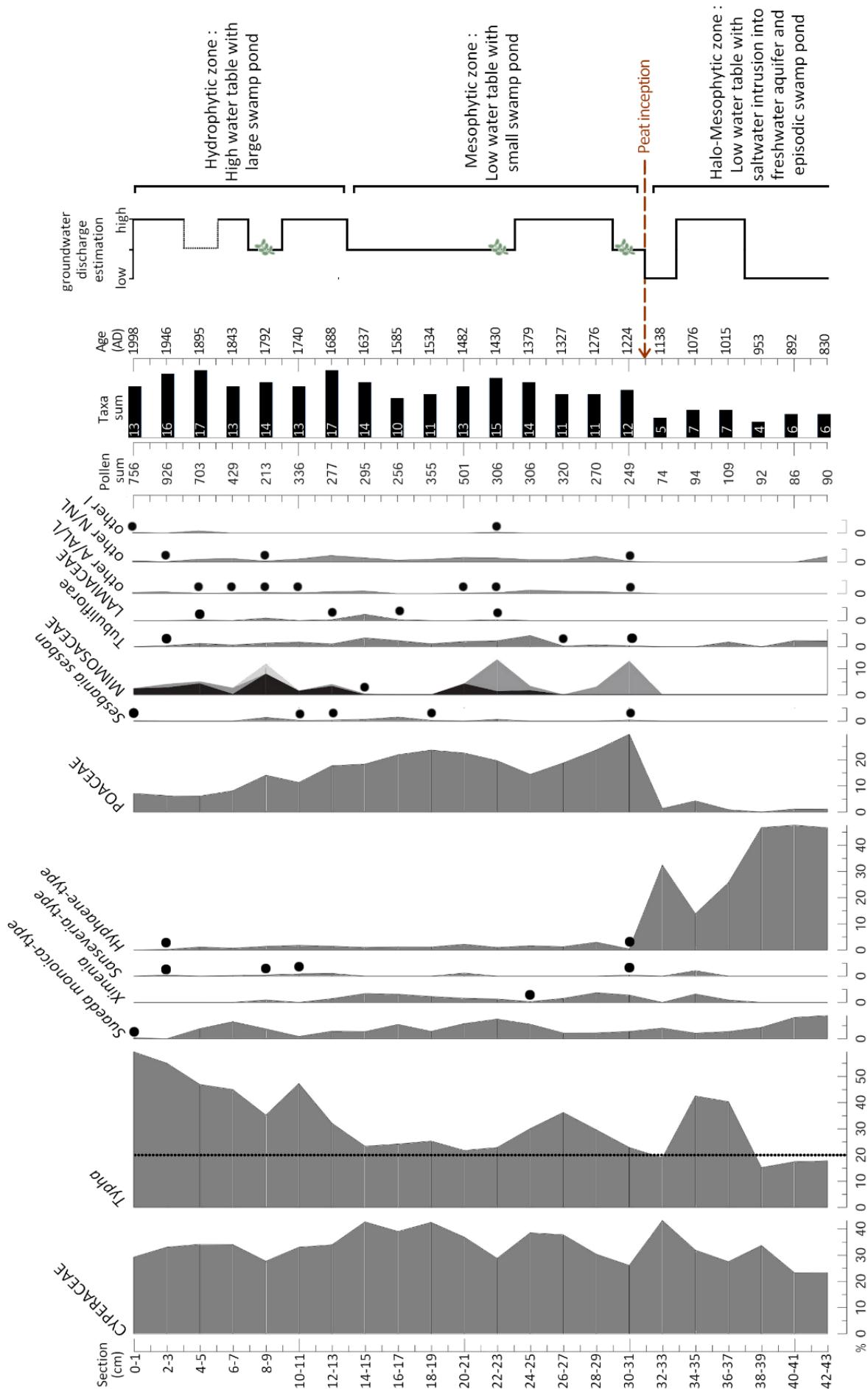
## 4.4. Results

### 4.4.1. $^{14}\text{C}$ dating and age models

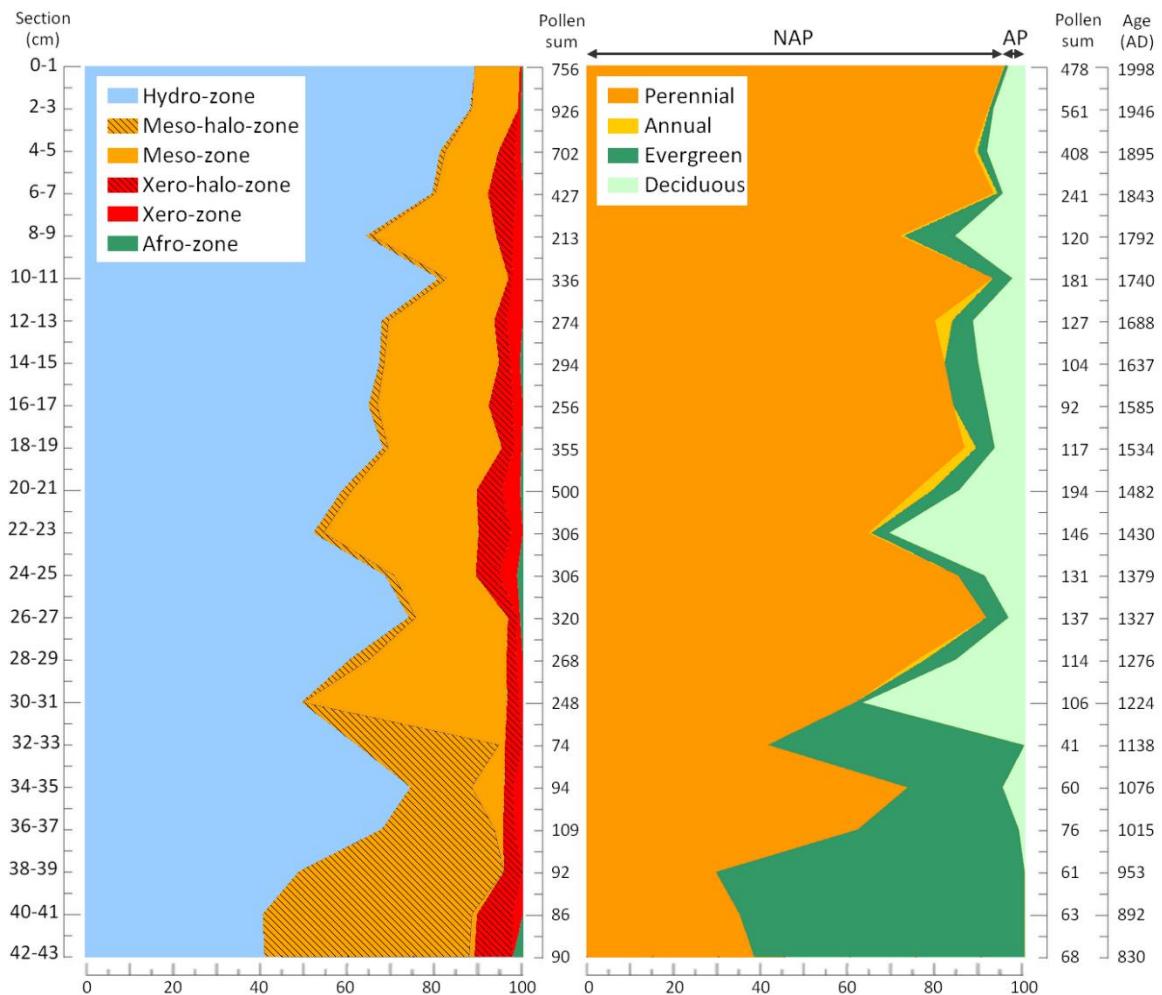
Four  $^{14}\text{C}$  dating of concentrated pollen grains were done at depths 21-22 cm and 30-31 cm in the upper part of the core, and at 37-38 cm and 42-43 cm in the lower part of the core (**Figure 4.4**). The ages obtained are expressed here as calibrated Anno Domini (AD): AD 1413 $\pm$ 24, AD 1261 $\pm$ 37, AD 884 $\pm$ 40 and AD 830 $\pm$ 158 (**Annexe table 4.2**). We chose to date pollen concentrates rather than bulk organic matter to minimize the reservoir effect (Li et al., 2012). Unable to say whether sedimentation was continuous between the upper and lower interval of the KN1A core, we made two independent age models and set the date of AD 2011 (year of core sampling) at 0 cm (**Figure 4.4** and **Annexe table 4.2**). For the upper interval of the core (0–31 cm) the model is very well constrained. With a sedimentation rate of 0.37 mm/yr, it respects the dating at 0 cm, 21-22 cm, and 30-31 cm. For the lower interval of the core (31–43 cm), there is more uncertainty. Considering the  $^{14}\text{C}$  dating errors, we calculated a mean sedimentation rate of 0.38 mm/yr as well as maximum (2 mm/yr) and minimum (0.2 mm/yr) sedimentation rates to define margins of error for the average sedimentation rate (**Figure 4.4** and **Annexe table 4.2**). Based on the sedimentary facies, the lower interval of the core is probably playa lake margin sediments. It is therefore possible that there are erosion surfaces and gaps in this part. Consequently, it is necessary to remain cautious with the age model of this lower core interval.

### 4.4.2. Pollen preservation and representation

Core KN1A produced an abundant pollen record with well-preserved pollen grains particularly in the upper part of the core (above 30-31 cm). In total, 46 different pollen types were identified (**Annexe table 4.1**). They represent various habitats, and changes in the paleovegetation between AD 830 to 2011 that are likely directly associated with changes in the spring flow and the level of groundwater table during the last 1200 years, and not directly with changes in the local climate (**Figure 4.5**). Taxa of the Afromontane, the xerophytic, and halo-xerophytic zones do not show any major changes in their relative abundances, which remain low throughout the record (<6%, <9%, and <2.5%, respectively) (**Figure 4.6**). We interpret this as the coring site namely the Afromontane forest, the *Acacia-Commiphora* bushland, and the salty lake margins.



**Figure 4.5** - Pollen diagram with major pollen types and interpretation of the pollen signal. Mimosaceae include *Acacia* in black, *Albizia*-type in grey and *Calpocalyx*-type *letestui* in light grey. Black dots indicate pollen percentages <0.5%. Groundwater discharge estimation is based on percentage of *Typha*, *Hyphaene*-type, and Mimosaceae. The leaf symbol represents peaks of Mimosaceae (*Acacia* or *Albizia*-type).



**Figure 4.6** - Synthetic pollen diagrams by ecological zones (undifferentiated excluded, cf., Table S1) and by life cycle/leaf phenology (undifferentiated excluded, cf., Table S1). AP: arboreal plants; Halo: halophyte; Hydro: hydrophyte; Meso: mesophyte; NAP: non-arboreal plants; Xero: xerophyte.

Pollen assemblage of the top core (0-1cm) is consistent with the vegetation surrounding the coring site in 2011. Pollen of *Typha* (59%) and *Cyperaceae* (29%) are abundant, typical of *Typha* sp. and *Cyperus* sp. swamps, but pollen of *Hyphaene* are not recorded, despite the presence of some palm trees about 50 m from the site (**Figure 4.2**). The small hydrophytic tree *Sesbania sesban* present at the site accounts for 0.1% of the total pollen sum. Thus, *Hyphaene* and *Sesbania sesban* are likely under-represented in the pollen assemblages. Poaceae in the surrounding short herbaceous swamp near the site only account for 7% of the total pollen sum. The pollen signal of allochthonous plants include *Suaeda monoica*-type (0.4%) from the salty lake margins, *Acacia* (2.4%) from the nearby (ca. 1 km away) tall *Acacia xanthophloea* woodland, and *Maerua*-type *crassifolia* (0.1%) from the surrounding xerophytic bushland.

#### **4.4.3. Three successive pollen zones indicate increasing groundwater influence since AD 1200**

*From the base of the core (43 cm) to the marked change in lithology (at 32-31cm).*

The lower part of the core, below 32 cm (AD 830 to AD 1140 is silty clay. The pollen assemblage is dominated by *Hyphaene* (>14%), *Typha* (15-43%), and *Cyperaceae* (23-43%) and taxa diversity is low (5 to 7 taxa at the most). This suggests that the site was not a wetland, but a dense *Hyphaene* palm woodland almost devoid of grasses within the halo-mesophytic zone, where vegetation has access to fresh groundwater while still being very much influenced by the salty water of Lake Eyasi. The presence of an *Hyphaene* palm woodland at the coring site at this time indicates the presence of a persistent halo-mesophytic zone (**Figure 4.3**).

*From 32-31 cm to the top.* The marked change in lithology observed between 32 and 31cm (dated AD 1138 and 1224) is also marked by a drastic change in the pollen composition reflecting a greater influence of groundwater at KN1 site after AD 1200 (above 32 cm). The upper part of the core, above 31 cm, is peat. It has been developing since AD 1200. In this part of the core, *Hyphaene* pollen represent <3% of the total pollen sum. Compared to the lower part of the core, *Poaceae*, *Typha*, *Cyperaceae*, and *Acacia* pollen are most abundant, and pollen taxa diversity is greatest (10 to 17 taxa) (**Figure 4.5**). The coring site which was covered by palm trees before 30-31 cm (AD 1200) was rapidly transformed afterwards into a grassy, species-rich woodland with trees, shrubs, and lianas with various species of *Acacia*, *Albizia*, *Celtis*, *Combretum*, *Erythrococca*, *Macaranga*, and *Ximenia*, among others (**Annexe table 4.1**). This vegetation change indicates that fresh groundwater had greater influence on vegetation than the salty water from Lake Eyasi after AD 1200 (**Figure 4.6**). Significant increases of *Mimosaceae* pollen at depths 31-30 cm (only *Albizia*), 23-22 cm (mainly *Albizia*), and 9-8 cm (*Acacia* and *Calpocalyx*) occur when less than 30% of *Cyperaceae* and slight increases of *Poaceae* are recorded, suggesting that a lowering of the water table favored the development of a grassy *Acacia* or *Albizia* woodland near the swamp (**Figure 4.5**). Finally, between 23-11 cm (AD 1430–1740), low and stable *Typha* percentage (25%), along with increased abundance of *Ximenia*, *Poaceae*, *Cyperaceae*, and *Tubuliflorae* and the absence of *Mimosaceae* suggests a period of low but constant groundwater discharge. The presence

of taxa such as *Sesbania* indicates that the area still includes stable freshwater outlets (**Figure 4.5**).

*From 14 cm to the top of the core.* *Typha* is a tall helophytic herb that grows in shallow, more or less stagnant water bodies where water is relatively fresh (not too alkaline or saline) (Agnew and Agnew, 1994; Greenway and Vesey-Fitzgeral, 1969). Its continued presence in the KN1A pollen record shows that stagnant freshwater wetlands are present in the vicinity of the coring site at least since AD 830 (the base of the core). It is interesting to note that plants of the hydrophytic zone are mainly represented by pollen of *Typha* and Cyperaceae (**Annexe table 4.1**), but the marked increase from 45% to >85% observed from the bottom to the top of the KN1A core (**Figure 4.6**) is almost exclusively due to *Typha*, whose relative abundance triples (from 18% to 59%) mainly after AD 1637 (at depth 14 cm) to the detriment of pollen types from the mesophytic zone (decrease from 45-50% to <10%). Hence, although the site became swampy by ca. AD 1224 (31 cm), it is only by AD 1637 (14 cm) that the *Typha* swamp began to expand, until reaching its actual size.

Groundwater discharge flow and, consequently, the size of the wetland thus appear to have progressively increased from AD 1600 to the present, where they are at their maximum. Spikes of *Typha* pollen, however, can be observed at several depths, with the most prominent ones at 34-37 cm, 10-11 cm, 4-5 cm, and 2-3 cm (**Figure 4.5**). They appear to reflect important but episodic expansions of swamps, likely as responses to important outflows in the groundwater discharge flow around AD ~1050, ~1330, ~1740, ~1840, and ~1946-2011 (**Figure 4.5**). Over the 1200 years represented by this pollen record, however, there is no pollen signal for a well-developed spring forest. This is attested throughout the pollen record by the abundance of Poaceae (up to 30%) and a low pollen taxa diversity, which never exceeds 20 taxa per sample and is essentially represented by herbaceous taxa.

## 4.5. Discussion

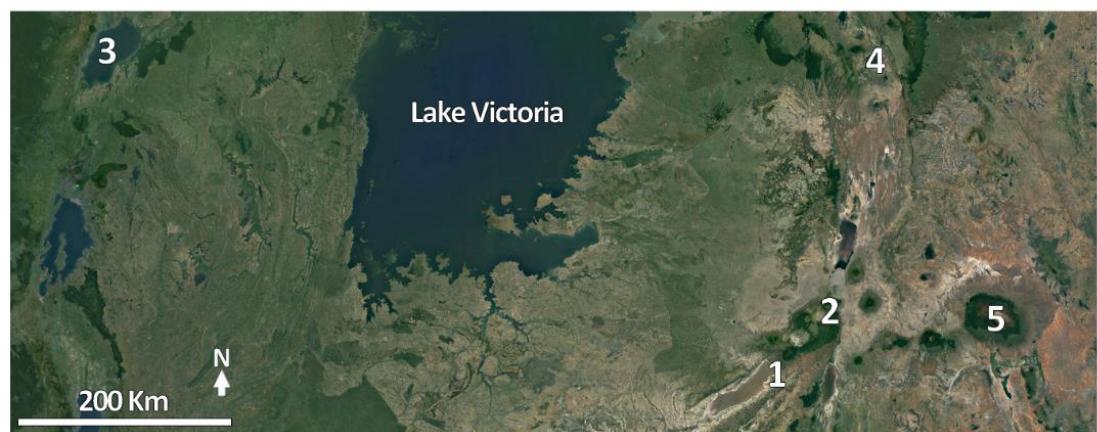
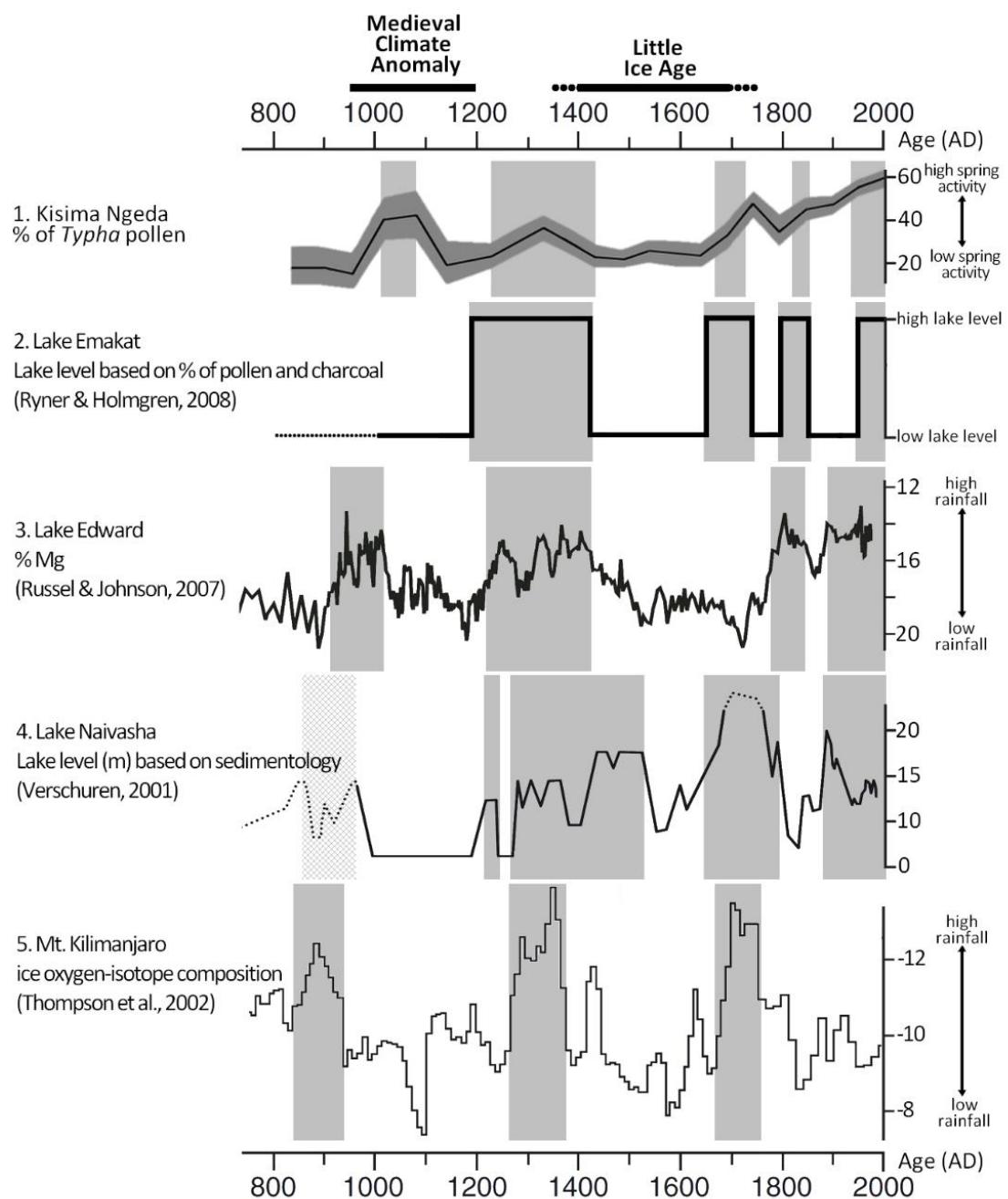
### 4.5.1. Microhabitats changes and groundwater level

KN1A core lithology and the pollen record indicate that a predominantly halomesophytic zone with *Hyphaene* palm woodland was present at the core site some 1200 years ago (from ~AD 830 to 1150), with possibly a short-lived wetland present around AD 1050 (which is imperceptible in the core lithology though) (Shilling, 2013). Such vegetation zonation suggests more saline/alkaline soil conditions before AD 1200 compared to the present, which could be due to a higher lake level than the current level, or a reduction in the flow and discharge of fresh groundwater. Based on our data, we can rule out the hypothesis of a higher lake level than the current level because, if Lake Eyasi were higher, then it is unlikely that pollen grains of *Hyphaene* palms would have been preserved in greater abundance than pollen of the salt-loving species (e.g. the Amaranthaceae *Suaeda monoica* and *Volkensinia prostrata*), which thrive on salty lake margins. The hypothesis of a reduced flow and discharge of fresh groundwater should be favored over that of a low lake level also because when pressure of fresh water on the salt water aquifer of the lake is low, salty water can penetrate inland (e.g., Fan et al., 1997), leading to more saline soil conditions and, therefore, to the expansion of the mesohalophytic palm zone.

Around AD 1150-1200, while an organic-rich peat started developing at the coring site, a dramatic change in the vegetation also occurred as the *Hyphaene* palm woodland was rapidly replaced by a swamp with a grassy woodland with abundant *Acacia*, rare *Hyphaene*, and many other species in the immediate surroundings, as shown by the increased pollen taxa diversity in this upper part of the core (**Figure 4.5**). The establishment of a spring outlet around AD 1150-1200 that still persists today is clearly linked to a higher water table and, consequently, to larger and more regular groundwater discharges than ever recorded before. The replacement of palm trees by grassy Mimosaceae woodland also attests for reduced soil salinity from around AD 1150-1200 onwards. After AD 1200 and until 2011, the KN1A pollen record shows no evidence that the vegetation near the coring site underwent significant changes due to new saline/alkaline water intrusions. This may be explained by the fact that if saline/alkaline water intrusions occurred, then they were not persistent enough as to trigger significant

changes in the vegetation, or that in this area of Lake Eyasi, saltwater intrusions into freshwater aquifer respond to changes in the fresh groundwater flow rather than to changes in the lake level. The analysis of diatoms preserved in KN1A core should provide more information regarding short-lived high lake events that may have occurred since the peat inception around AD 1150-1200 (Shilling, 2013).

From the peat inception (AD ~1150-1200) to AD 1300, the groundwater discharge as inferred from the relative abundance of *Typha* pollen increased. After a peak at AD ~1330, it then dropped slightly and stayed stable until AD ~1640. The spring activity inferred from the percentage of *Typha* increased again after AD ~1640, and three peaks of groundwater discharge occurred at AD ~1740, ~1840, and ~1946-2011 (**Figure 4.7**). Highest spring activity seems to occur today, as the percentage of *Typha* pollen has never been so high throughout the record.



**Figure 4.7** - Spring activity of Kisima Ngeda based on pollen percentages of *Typha* over the past 1200 years with 95% confidence interval (dark gray envelope) compared with other regional proxy records reflecting moisture variations. Light gray stripes indicate the "wet" phases.

#### **4.5.2. Comparison with other regional records**

An increase of fresh groundwater flow and discharge is expected after increased rainfall in the highlands (Cuthbert and Ashley, 2014). It is therefore possible to compare KN pollen record with other regional records of past hydroclimate.

Before AD 1200, a generally dry interval of ~400 years is recorded at Lake Emakat (partially) (Ryner et al., 2008), Lake Duluti (Öberg et al., 2013), Lake Naivasha (Verschuren, 2001), Lake Edward (Russell and Johnson, 2005), and is also observed in the spring activity of Kisima Ngueda (as inferred from the percentage of *Typha* pollen in KN1A core). We note, however, that this dry period seems to be interrupted by a ~50-100 years-long wet episode around AD 1050 in several of these records (**Figure 4.7**). The peat inception at Kisima Ngueda at AD ~1150-1200 is concomitant with the beginning of a wet (pluvial) episode at L. Emakat, L. Edward, L. Challa, and L. Naivasha around AD 1200 (Russell and Johnson, 2007; Ryner et al., 2008; Tierney et al., 2013; Verschuren et al., 2009), and with peat inceptions at other groundwater-fed localities in Kenya such as the Solai (Goman et al., 2017) and Loboi swamps (Ashley et al., 2004; Driese et al., 2004). Altogether this seems to mark the end of a generally dry period well recognized in the northern hemisphere as the (warmer than normal) Medieval Climate Anomaly (MCA). In Africa, the MCA (with a core period of AD 1000–1200) is characterized by important changes in the hydroclimate, with strong regional differences. The latest data compilation shows that Eastern Africa was drier south of the equator and west of a SSW-NNE line separating the drier inland sites (L. Edward, L. Victoria, L. Masoko, L. Tanganyika, Kisima Ngueda included) from the wetter coastal eastern African sites during the MCA (L. Challa, L. Malawi) (Lüning et al., 2018).

From AD 1200 to ~1400-1450, the development of a peat and a grassy Mimosaceae woodland at Kisima Ngueda occurs while all neighboring lakes record high levels (Russell and Johnson, 2007; Ryner et al., 2008; Verschuren, 2004, 2001) and high rainfall is inferred at Kilimanjaro (Thompson et al., 2002). Between AD 1400-1600, the groundwater discharge at Kisima Ngueda is low but sufficient to preserve the wetland bordered by grassy woodland. This interval corresponds to low lake level at L. Emakat and drought inferred at L. Edward (Russell and Johnson, 2007) and at other small lakes in western Uganda (Russell et al., 2007) and Malawi (Brown and Johnson, 2005; Johnson et

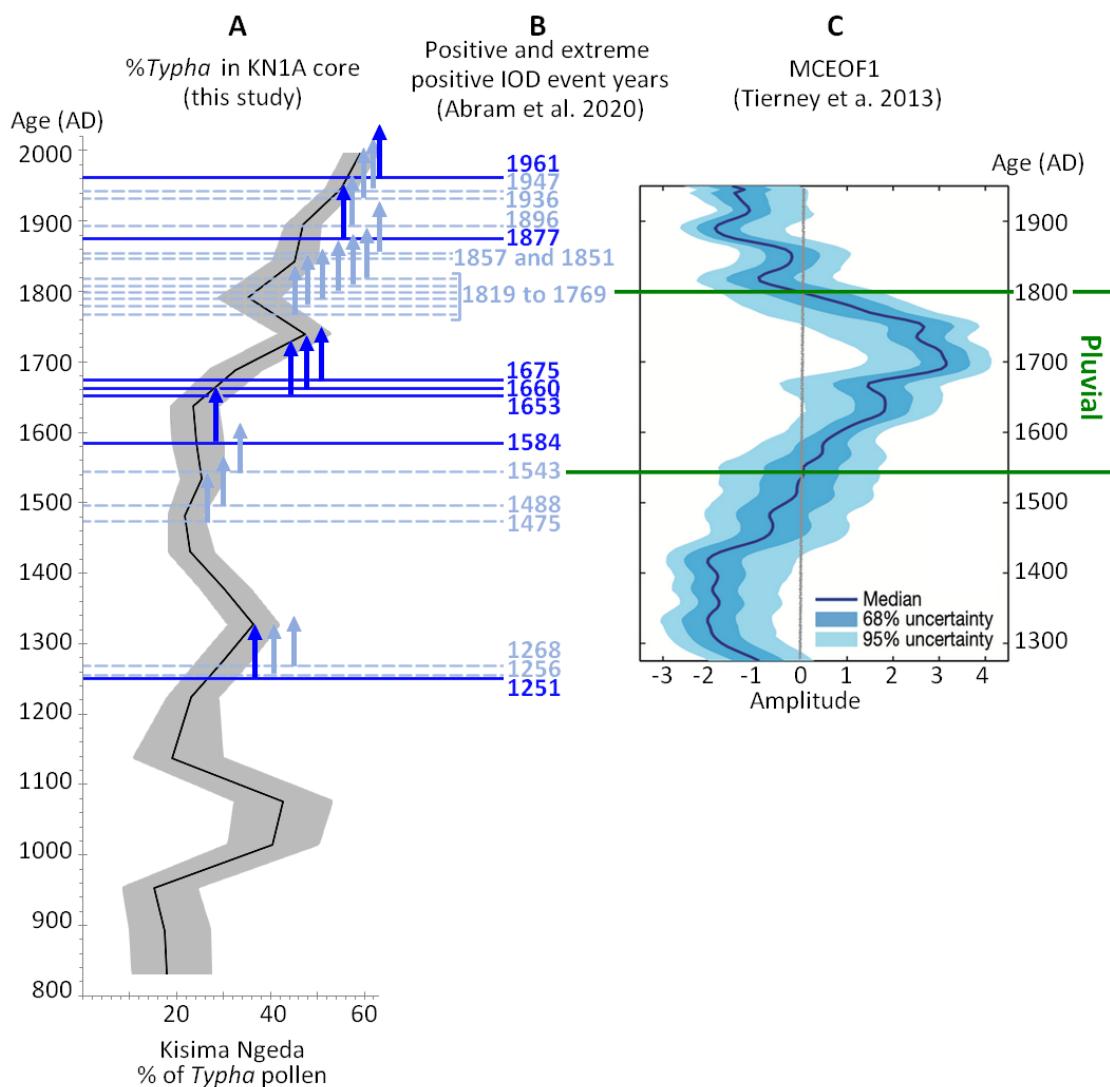
al., 2001; Johnson and McCave, 2008), but to high lake level at L. Victoria (Stager et al., 2005) and L. Naivasha (Verschuren, 2001; Verschuren et al., 2000). Lake Challa, L. Masoko and L. Tanganika, do not match any of the pattern cited before, and show contrasting and varying patterns over this 1400-1600 AD time period (Tierney et al., 2013). Hence, at the scale of Eastern Africa the Little Ice Age, which seems to be recorded from as early as AD 1350 to as late as AD 1750 (**Annexe table 4.3**), shows very contrasting patterns. At Kisima Ngeda, the LIA doesn't appear well-marked in the pollen record; yet, the spring activity was higher between AD 1400-1700 than during the Medieval Climate Anomaly. Later on, between 1650-1800, while drought is recorded at L. Malawi, L. Masoko, L. Tanganika, and L. Edward (i.e., in the western rift zone), a pluvial interval is recorded at L. Challa, L. Naivasha and L. Emakat (in the eastern rift zone). From AD ~1790 spring activity has been increasing at Kisima Ngeda and has not weakened since (**Figure 4.7**), despite marked droughts recorded throughout the African continent in the 1820s and 1830s (Nicholson et al., 2012) and since 1990 (Lyon and DeWitt, 2012).

As shown by the regional comparison attempted here and more thoroughly elsewhere, trends between different site archives are rarely shared over the last 1200 years (i.e., over the entire period) in Eastern Africa (Nash et al., 2016; Tierney et al., 2013). The Kisima Ngeda record of spring activity over this period is sometimes consistent with sites in the south and west (e.g., Masoko, Tanganika, Edward) and sometimes with the more coastal sites (e.g., Challa and Naivasha) depending on the time interval considered, but it also shows no signs of any early-19th century drought, and no signs either for other severe drought events in the final decades of the 19th century like in the Kenya Rift Valley, or in Ethiopia or elsewhere in northern Tanzania where as much as 40-75% of pastoralist Maasai may have succumbed (Iliffe, 1987). The climate mitigating effect of the Kisima Ngeda spring needs to be examined in more details.

#### **4.5.3. Spring vegetation response to IOD positive events**

It is now established that precipitation in Eastern Africa is strongly correlated with sea surface temperatures (SSTs) of the Indian Ocean (Latif et al., 1999; Tierney et al., 2013), and that the Indian Ocean Dipole (IOD) is responsible for extreme precipitations in Eastern Africa and coeval extreme droughts in Indonesia (Saji et al., 1999). IOD is an air-sea interaction process independent of El Niño/Southern Oscillation which is characterized by an anomalous west to east gradient of decreasing (instead of increasing)

SSTs as well as anomalous winds and sea heights in the tropical Indian ocean (Saji et al., 1999; Webster et al., 1999). According to the  $\delta^{18}\text{O}$  record in the corals from the southern Mentawai Island chain, offshore of Sumatra, several positive and extreme positive IOD events occurred during the last millennium (Abram et al., 2020). **Figure 4.8** shows that peaks of *Typha* pollen, which we interpreted as wetland expansion following increased groundwater flow and discharge, systematically occurred  $\sim 75$  years after each extreme-positive and positive IOD events. The correspondence between spring activity and IOD is striking. Positive IOD events may also be observed on the Kilimanjaro  $\delta^{18}\text{O}$  ice record (Thompson et al., 2002), if accepting some uncertainty associated with the age model (**Figure 4.7**).



**Figure 4.8** - Kisima Ngeda spring response (wetland expansions) to mega-regional and regional events. **(a)** Relative abundance of *Typha* pollen in KN1A core at Kisima Ngeda, Lake Eyasi, N Tanzania. Brown arrow indicates peat inception. **(b)** Positive (light blue) and extreme positive Indian Ocean Dipole (IOD) events (dark blue) according to the coral  $\delta^{18}\text{O}$  record of the southern Mentawai Islands, offshore of Sumatra (Abram et al., 2020). Arrows represent 75 years after positive IOD events. Note that, to date, there is no isotopic record of IOD events older than AD 1240. **(c)** Amplitude of the first Monte Carlo empirical orthogonal factor (MCEOF) showing a coherent statistical signal of rainfall (as inferred by various proxies) in Eastern Africa over the AD 1300-2000 time period. Kisima Ngeda spring activity exhibits a different signal.

At present and over the 1200 year-long period represented by KN1A record there is no evidence for a well-developed spring forest like in the northern edge of saline/alkaline Lake Manyara (**Figure 1**) (Greenway and Vesey-Fitzgerald, 1969; P. E. Loth and Prins, 1986). In Manyara spring forest, arboreal pollen taxa diversity in surface samples is up to 27 taxa/sample, and mirrors the tree and shrub diversity of the spring forest (Barboni et al. 2019). At Kisima Ngeda, arboreal pollen taxa diversity never exceeds 17 taxa/sample in surface soil nor in the fossil (core) samples (Barboni et al., 2019) (**Figure 4.5**). Low taxa diversity in KN1A core is not related to preservation (except in the bottom part of the core) since it's also observed in the modern soil samples (Barboni et al., 2019). Forests with stratified canopy develop when hydrological conditions are favorable for the growth of highly-demanding large-leaved evergreen tree species, in particular when amount of water is not limiting and access to water is stable for several decades or even centuries. The KN1A core shows that although groundwater input has been stable, even increasing in the last 1250 years, a multi-story evergreen spring forest never developed at Kisima Ngeda. This can be explained by several factors. (1) The groundwater flux at Kisima Ngeda (south-eastern foothills) is likely lower than that at Manyara (south-western foothills), because the south-western slopes of Ngorongoro Volcanic Highlands receive less rainfall than the south-eastern slopes (Herlocker and Dirschl, 1972). Difference in rainfall between the two areas is so great that the lower edge of the Afromontane forest is at *ca.* 1600 m on the south-western slope (Lake Manyara Basin) and as high as 2450 m on the south-eastern slope (Lake Eyasi Basin) (Herlocker and Dirschl, 1972; Norton, 2019). (2) The vegetation in these spring systems also depends on the geomorphology of the drainage basins, which influences water flows and drainage. At Gorofani, located a few kilometers east of Kisima Ngeda, tall and large trees of *Ficus* sp., *Tamarindus indica* and *Cordia* sp. are still present despite heavy grazing and human presence. This suggests that the local setting might have been more favorable than in KN for the development of a multi-story groundwater forest like at Manyara (D. Barboni, personal observation). (3) The groundwater pH and conductivity are very different between Kisima Ngeda and Manyara spring wetlands. The pH is more acidic at Manyara (pH 5.7) than Kisima (pH 7.4-7.6), and conductivity (dissolved elements) is low and consistent at Manyara (354-384 µS), but high and highly variable at Kisima (640-5600 µS). Water temperature is similar at both sites (about 21°C) (Ashley, unpublished data).

#### **4.5.5. Kisima Ngeda, a dual and therefore sustainable groundwater system?**

The KN1A pollen record shows that during low spring activity the groundwater table at Kisima Ngeda is still high enough to sustain a palm woodland (e.g. before AD 1200 when there is a salty influence) or a grassy Mimosaceae woodland. The hydrological system that feeds this densely wooded localized vegetation belt in the northeastern margin of Lake Eyasi remains active even during arid periods of ~300 years during the Medieval Climate Anomaly, and doesn't seem to be influenced by dry climatic phases. Yet, the pollen record also shows that, at the same time, the groundwater flow at Kisima Ngeda responds very quickly (on a decadal scale) to wet climatic events via discharge peaks. A groundwater system cannot be at the same time non-sensitive and very sensitive to climatic variations. Consequently, the springs around Kisima Ngeda are probably fed by two independent groundwater systems: a groundwater with temporally sensitive short groundwater response time (GSS) and a groundwater with temporally insensitive long groundwater response time (GTI) (Cuthbert et al., 2019). The GSSs are close to the surface with short (decennial) response time, while the GTIs are deeper with large response time, making altogether a dual, sustainable system (Cuthbert et al., 2019).

Additional geological and hydrological data will further enhance our understanding of the Ngorongoro area groundwater GSS and GTI systems (Norton and Ashley, unpublished data). Yet, what we show here is that in this region, some spring systems such as Kisima Ngeda are sustainable over more than a millennium due to stable groundwater discharge, but do also respond to punctual discharge linked to decennial or centennial climatic variations. Such system, therefore, could have sustained early hominin activity such as the active exploitation of spring as observed at Olduvai (Ashley et al., 2009), which could not have developed if springs were ephemeral. Given the reconstructed low precipitation amounts during the Pleistocene in Olduvai area (250-700 mm/yr), it is likely that surface water (permanent rivers) was limited, making early hominins highly dependent on stable freshwater sources (Magill et al., 2013a). In agreement with Magill et al. (2016b), our study confirms the stability of these systems on a time scale compatible with human activities and reinforces current paradigms. Our study also supports the idea that this type of groundwater system is consistent with the development of stable woodlands that may have facilitated the movement of the early hominins in the Eastern African rift during arid climate phases (Cuthbert et al., 2017).

#### **4.5.6. Impact of springs on the geological and paleobotanical records at hominin sites**

As mentioned, the hydrological system of Kisima Ngieda seemed to be a potential analog to some Olduvai sites because of the presence of a (paleo) lake nearby that is saline/alkaline, of palms, and of (paleo) springs with fresh groundwater whose recharge occurs in the Ngorongoro Highlands. The Olduvai pollen record (Bonnefille, 1984), however, differs from the KN1A pollen record by the abundance of Poaceae (35-75%), rare *Typha* (0-10%), the absence of *Hyphaene* or other palm taxa, and the abundance of pollen markers of the *Acacia-Commiphora* bushland. Olduvai pollen data alone suggest open semi-desert environment comparable to the present-day vegetation with the presence of very reduced or distant fresh-water swamp, but no spring forest or woodland. Yet, other vegetation proxies ( $\delta^{13}\text{C}$  on pedogenic carbonates, phytoliths, macro-remains) indicate highly heterogeneous woody cover, abundant palms, and *Typha* wetlands (Arráiz et al., 2017; Ashley et al., 2010a; Bamford, 2012; Barboni, 2014; Barboni et al., 2010; Cerling and Hay, 1986; Magill et al., 2013b; Sikes and Ashley, 2007), while tufa mounts are geological evidences attesting for the presence of groundwater discharge areas (springs) at Olduvai (Ashley et al., 2010a). The absence of "spring indicators" in the Olduvai pollen record despite evidences in the geology, geochemistry and phytoliths and macro-remains may be explained by the low pollen production (e.g. *Sesbania sesban*) or low dispersion capacity (e.g. *Hyphaene* or *Acacia*) of spring plant taxa, which makes them less easily recordable than e.g. grasses. Pollen grains of *Hyphaene*, however, were recorded in the Pliocene of Laetoli (Bonnefille and Riollet, 1987). Hence, the depositional context is another factor that could explain the absence of spring pollen markers at Olduvai. As shown in this study, groundwater discharge areas trigger and maintain a landscape heterogeneity, not only in terms of vegetation by allowing the development of dense and tall woodlands and wetlands within an otherwise desert-like vegetation, but also in terms of depositional environments. Indeed, following the groundwater discharge and level, a given area may alternatively accumulate organic-rich peatland material (which favor pollen preservation) or become a sandy clay-rich soil (less likely to preserve pollen). The water table level and spatial position of the discharge areas (spring outlets) likely change over millennia, and swamps, woodlands, and palm groves (the spring-associated vegetation zones) are then displaced laterally by few kilometers. Yet, if these local

movements have potentially little impact on the wildlife, they may considerably disturb the geological and paleobotanical records.

Groundwater forests and woodlands have a very distinct but also very local pollen signal, which can go unnoticed (Barboni et al., 2019). Depending on the location of the sampling site, the period studied and the resolution of the record, this type of spring system can give a very heterogeneous vegetation signal, often interpreted as mosaic environment with swamps, groundwater woodlands, palm groves, bushland/wooded bushland, etc. Hence, to characterize hominin habitats in the East African Rift where springs are common, it is necessary to consider spatial and/or temporal resolution of paleoenvironmental reconstructions (Behrensmeyer and Reed (2013), and to search for evidences of paleo-springs in both the geological and the biological records (Barboni et al., 2019) to avoid interpretation biases.

## 4.6. Conclusions

The pollen analysis of the 43 cm-long core KN1A at Kisima Ngeda locality in the north-eastern shore of Lake Eyasi provided interesting insights into a dynamic groundwater system in the EARS during the last 1250 years. It shows that KN spring has been active at least since the last 1250 years, that the water table was low before 1200 AD (with discontinuous and sparse discharge at the ground level). It abruptly increased at 1200 AD, i.e. shortly after the so-called dry Medieval Climate Period, and shows a steady increase since 1672 AD. The groundwater table at Kisima Ngeda varied at the pace of regional precipitation changes that affected a large area encompassing Mt. Kilimanjaro, L. Naivasha, L. Emakat, and as far east as L. Edward. Yet, Kisima Ngeda spring system has persisted despite a 250 years-long dry period from 1400 AD to 1650 AD, in phase with the northern hemisphere Little Ice Age. Kisima Ngeda spring, therefore, is likely controlled by a dual groundwater system which allows maintaining wooded habitats even during dry phases, and which favors the development of wetlands following discharge peaks during wet phases. Such spring-fed wetlands can precisely record paleoprecipitation changes. Yet, because changes in the groundwater discharge modify the position (but not the presence) of different ecological zones in the spring area, the geological and paleobotanical record (from sedimentary cores) may have hiatuses or be

discontinued. The KN1A pollen record did not match the Olduvai fossil pollen record, despite obvious analogy in the geological, geomorphological and geochemical contexts, likely because spring plant taxa such as *Sesbania* and *Hyphaene* are poor pollen producers and dispersers. Although tufas are solid evidences of paleo-springs, they are not sufficient to infer the type of habitat and vegetation associated to it, because the density, development and diversity of the spring vegetation likely depend on the orientation and geomorphology of the drainage basins, and groundwater quality.

#### **4.7. Acknowledgments**

This work was supported by Aix-Marseille University (doctoral fellowship to BB), the Agence Nationale de la Recherche (HADoC Project to DB, ANR-17-CE31-0010-02), and a Waitt Grant/ National Geographic to GM Ashley (Grant # 168-11). Some  $^{14}\text{C}$  dates were determined by accelerator mass spectrometry at Artemis facility, UMS LMC14 Saclay, France, in the framework of INSU national service (grant to DB). We thank Jonathan Coutaz (CEREGE) for his assistance on the hydrogeological interpretations of this study.

## **5. Reconnaissance automatisée de grains de pollen modernes, fossiles, intacts et endommagés par réseaux neuronaux convolutifs multiples**

- Titre de l'article : Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains
- Auteurs : Benjamin Bourel<sup>1</sup>, Ross Marchant<sup>1</sup>, Thibault de Garidel-Thoron<sup>1</sup>, Martin Tetard<sup>1</sup>, Doris Barboni<sup>1</sup>, Yves Gally<sup>1</sup>, Luc Beaufort<sup>1</sup>  
<sup>1</sup>CEREGE, Aix Marseille Université, CNRS, IRD, INRAE, Coll. France, Technopole Arbois, 13545 Aix en Provence cedex 4, France
- Revue : *Computers and Geosciences* (<https://doi.org/10.1016/j.cageo.2020.104498>)
- Résumé : Pollen grains are valuable paleoclimate and paleovegetation proxies which require extensive knowledge of morphotypes and long acquisition time under the microscope. The abundance of damaged, folded, and broken pollen grains in the fossil register and sometimes also in modern soil and sediment samples, has so far prevented automation of pollen identification. Recent improvements in machine learning, however, have allowed reconsidering this approach. Here we present an automated approach which is capable of assisting palynologists with poorly preserved pollen samples. Called multi-CNNs, this approach is based on multiple convolutional neural networks (CNNs) integrated in a decision tree system. To test it, we built a system designed for three botanical families very common in the modern and fossil pollen assemblages of Eastern Africa, namely Amaranthaceae, Poaceae, and Cyperaceae. Our system was tested on stacked optical images of 8 pollen types (6 Amaranthaceae, 1 Poaceae, 1 Cyperaceae) using a training dataset of 1102 intact pollen grains and three validation datasets of intact (276 grains), damaged (223 grains), and fossil pollen (97 grains). We show that our system successfully recognizes intact, damaged, and fossil pollen grains with very low misclassification rates of 0%, 2.8%, and 3.7%, respectively. The use of augmentation on stacked optical images during the training increases classification accuracy. Following a palynologist's approach, our system allows grains without obvious characters to be classified into a class of high taxonomic level or as indeterminable pollen. This is the first

software able to process grains with a wide range of taphonomical stages, which makes it the first truly applicable to automated pollen identification of fossil material.

- Mots clefs : damaged pollen, fossil pollen, machine learning, image analysis, z-stacking, Amaranthaceae

## 5.1. Introduction

Since the late 1960s, the palynologist community has been interested in the automation of pollen grain recognition (Flenley, 1968). This desire comes from a real need to reduce the long and tedious process of pollen grain identification and counting, which requires an extended knowledge of palynofloras in order to obtain accurate and statistically valid paleovegetation inferences. Pollen counting and recognition process is an important limiting factor that makes it difficult to build large databases. To date, with the advances of large datasets for paleoenvironmental studies to resolve high spatial and/or temporal resolution, there is a crucial need to reduce the time involved in pollen data acquisition. Palynology is also a discipline at risk today because acquiring sufficient knowledge of palynofloras requires years of experience, particularly in tropical regions where plant diversity is highest. Despite the fact that pollen grains are still the only paleovegetation proxies allowing plant identifications up to the genus and species levels, their study is easily supplemented by other indicators such as  $\delta^{13}\text{C}$ . They require less physical presence of scientists in the acquisition although they do not allow such fine environmental reconstitutions.

In the East African Rift, many paleontological areas have preserved abundant fauna, including hominin remains of Pliocene and Pleistocene age, but few have preserved paleobotanical remains. Yet, pioneering pollen studies in the early 70s have proved to be a powerful tool for inferring paleoenvironments associated with early hominins in Africa (Bonnefille, 1979). Among these, the *Australopithecus afarensis*-bearing Hadar formation in the North Ethiopian Rift was extensively studied, and several levels proved to be rich in pollen grains (Bonnefille, 2010; Bonnefille et al., 2004; Bonnefille and Riollet, 1987). Those levels were analyzed by palynologists of Bonnefille's team, usually over several weeks, in order to obtain adequate pollen counts (generally between 300 and 1000 grains

depending on the pollen diversity of each sample) as the depositional and arid climate contexts were not favorable to pollen preservation (R. Bonnefille, pers. Com.). Recently, the Hominin Sites and Paleolakes Drilling Project (HSPDP) drilled in this area to obtain paleoenvironmental records with high temporal resolution of the variations of the local hominin environment between 2.9 and 3.6 million years (Ma) ago (Cohen et al., 2016). We have been studying two of those HSPDP cores, collected from near the lacustrine depocenter of Hadar Basin, the Northern Awash Osi-Isi (NAO, 11.31518° N, 40.73689° E) and from a location 3 km east of NAO, the Northern Awash Woranso (NAW, 11.32535° N, 40.76491° E). NAO and NAW were cored respectively to a depth of 187 and 245 m below surface (Mohan et al., 2016). Analysis of 8 samples out of 175 from these two cores took one of the authors (B.B.) several months to complete, because both the preservation and concentration of pollen grains were low. A high-resolution pollen record at this site would give a fantastic record of paleovegetation changes during the time when *Australopithecus afarensis* thrived and when the first representatives of the *Homo* genus emerged (Maslin et al., 2015; Villmoare et al., 2015), but such a study would require more than one scientist working full time over several months. It is a classic case of “arid” sedimentary archives with very important environmental pollen data that cannot be correctly studied due to human limitations. Testing and exploring the potential of automated recognition to pollen taxa is, therefore, crucial.

In the 1960s, computer performance, image processing, and machine learning techniques were not sufficiently developed to realize Flenley (1968)’s visionary concepts of pollen recognition. During the last decade, however, dramatic improvements opened the way to automated recognition of pollen, through different approaches such as (1) feature extraction, (2) texture analysis, and (3) the combination of both. (1) Regarding feature extraction, the first breakthrough study by García et al. (2012) combined morphological methods based on contour feature extraction of light microscopy (LM) images with a hidden Markov model (HMM) classifier. This approach allowed classifying 17 different pollen types from 11 different botanical families of honey bee’s plants with 98.8% accuracy. Kaya et al. (2013) used morphological measurements (polar axis length, equatorial axis length, colpus weight, exine thickness, etc.) with a data mining classifier to predict group membership for data instances. This approach allowed 20 very similar species of pollen grains in the genus *Onopordum* to be classified with an accuracy of 90.6%. The work is impressive but requires an initial manual/subjective definition of the

recognition characteristic features and is only applicable to intact pollen grains. In addition, the accuracy of Kaya et al. (2013)'s model is dependent of the choice of characteristic features (Treloar et al., 2004). (2) Another set of studies focused on texture analysis, with texture features derived from gray level patterns of the grain surface (Fernandez-Delgado et al., 2003). Using a combination of local linear transformation (LLT) feature vectors and a support vector machine (SVM) classifier, Fernandez-Delgado et al. (2003) were able to classify LM images of fresh pollen of five pollen types (in four different families) with an accuracy of 76%. Later, a more complex approach was developed by Marcos et al. (2015) using a texture feature vector composed of gray-level co-occurrence matrices (GLCM), log-Gabor filters (LGF), local binary patterns (LBP) and discrete Tchebichef moments (DTM), and Fisher's discriminant analysis (FDA) and k-nearest neighbor (KNN) to reach a final classification at the family level. This study achieved an accuracy of 94.8% for LM images of fresh pollen grains of 15 pollen types in 12 families (Marcos et al., 2015), but the texture approach has been shown to be sensitive to grain rotation, lighting, image noise and the quality of texture preservation (Han and Xie, 2018). To circumvent these problems, Han and Xie (2018) proposed to use local decimal patterns (LDP) instead of LBP, an approach robust to the noise and rotation of pollen images but which, nevertheless, could not tackle the challenge of recognizing damaged pollen.

(3) A combination of texture features and morphological analyses represents the last development. Rodriguez-Damian et al. (2006) developed a classification system using both shape and texture analysis of LM images of pollen grains directly extracted from plants. Using a minimum distance classifier (MDC), multilayer perceptron (MLP) and SVM, three morphologically similar pollen types of Urticaceae were classified with an accuracy of 89%. Later, Chudyk et al. (2015) used a combination of 15 morphological features and five textures features on a set of LM images including one pollen type of Poaceae, one of Asteroideae and three of Betulaceae. By using SVM and Random Forest classifiers, they reached a success rate of 89% for intact pollen grains. However, this approach still performed poorly on damaged pollen.

Recently, the development of convolutional neural networks (CNNs) has opened a new era in image recognition. CNNs automatically extract the features of interest in an image that are relevant to the classification task (Daood, 2018). With this approach, intact

pollen grains were identified with an accuracy of 97.2% with a LM images dataset (Gonçalves et al., 2016). CNNs with different topology were also tested and allowed classifying 23 pollen types with large morphological differences (Sevillano and Aznarte, 2018). Daood et al. (2018) recognized z-stacked LM images of intact pollen by combining CNN feature vectors with recurrent neural networks (RNN). They achieved an accuracy of 100% for the classification of 10 pollen types of different families. Comparisons of these methods and their success rates on the same data set of images showed that the approaches based on CNNs are always the most accurate, and point toward the efficiency of CNNs in classifying natural objects such as pollen.

All these approaches proved to be efficient only on modern and generally intact pollen grains, far from the reality faced by palynologists who must deal with damaged pollen due to taphonomic processes. Holt et al. (2011), Holt and Bennett (2014) and recently Khanzhina et al. (2018) noticed the absence of a system adapted to “broken, deformed and clumped pollen” and explained it by technological limitations. Han and Xie (2018) pioneered the analysis on damaged pollen grains, unfortunately their approach did not succeed. Punyasena et al. (2012), however, successfully classified fossil grains of two *Picea* species (*P. mariana* and *P. glauca*) with an accuracy of 93.8%. They used a KNN with layered learning systems based on kernel density, applied to extracted shape and texture features from 3D fluorescence microscope (FM) images. Using the same set of pollen grains as Punyasena et al. (2012), Kong et al. (2016) also obtained a high accuracy of 86.1% in classifying fossil grains of three *Picea* species (*P. mariana*, *P. glauca* and *P. critchfieldii*) using dictionary learning sparse coding (DLSC) on shape and texture features extracted from confocal FM images. However, the performances were largely dependent on the selection and quality of sample blocks (Han and Xie, 2018; Kong et al., 2016). These results are nonetheless very promising, but we note that the deterioration of the *Picea* pollen considered in those two studies is very subtle (as is generally observed in peat sediments, eg., Curry, (2007)), and far from resembling the Pliocene pollen records of Eastern Africa.

The aim of this study is to develop a new system of automated identification of pollen that could be applied to down core fossil pollen grains. This system should be able to handle damaged and broken pollen grains, and should be easily applicable to images obtained with light microscopy (LM), as is commonly done in palynological laboratories.

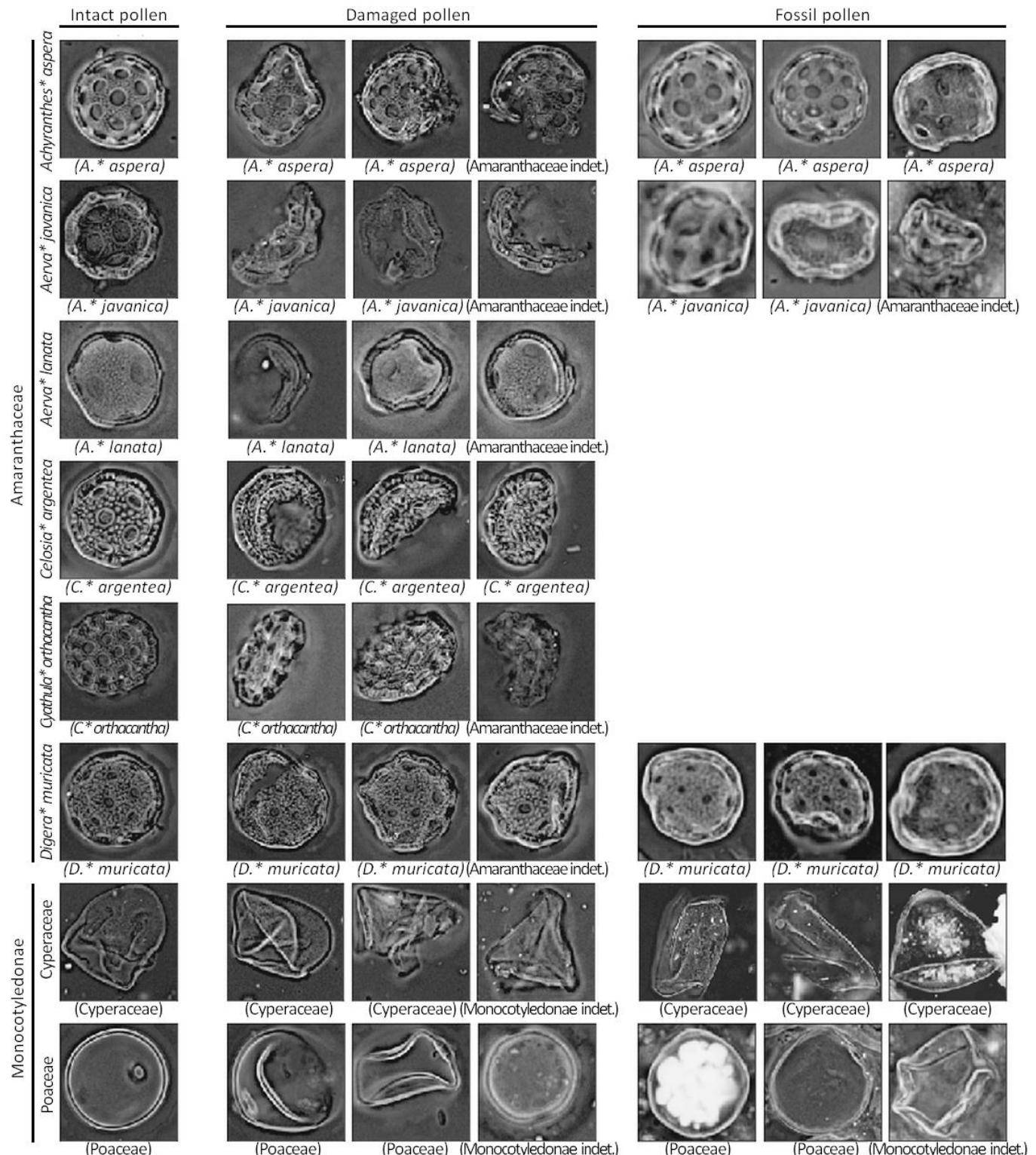
It should also allow identifications at higher taxonomical levels when identification at the species level is not reached.

Following recent developments in machine learning we chose to develop our system using sequential multiple CNNs. As a proof of concept for this study we chose eight pollen types: six among the Amaranthaceae, as they have very similar overall shape (spherical, periporate), one Poaceae (undifferentiated taxa), and one Cyperaceae (undifferentiated taxa), that is two Monocotyledonae with heterogeneous and varying surface patterns (psilate to finely verrucate). We focused on Amaranthaceae, Poaceae, and Cyperaceae because these three families represent 50%–90% of the total pollen assemblage in both terrestrial and marine archives from Eastern Africa (Bonnefille et al., 1987; Bonnefille, 2010). Pollen-based biome reconstruction for Africa also largely rely on the presence and relative abundance of Amaranthaceae pollen (Sobol and Finkelstein, 2018). Amaranthaceae species, despite exhibiting very similar pollen morphologies are important ecological markers (“African Plant Database,” 2019). Being able to identify and count separately different Amaranthaceae could definitely improve paleoenvironment inferences. As a first step, this study makes no attempts to separate types among Poaceae and Cyperaceae, also because such discriminations have only been achieved so far using Scanning Electron Microscopy (Mander et al., 2014).

## 5.2. Material and Method

### 5.2.1. Datasets and image acquisition

We prepared one training and three validation datasets by taking pictures of pollen grains under light microscopy (LM) (**Annexe table 5.1**). The training dataset consists of 1102 globally intact pollen grains (hereafter referred to as the training dataset). The three validation datasets consist of 1) 276 grains of intact pollen (hereafter referred to as the intact dataset), 2) 223 deformed, folded, fragmented and/or broken grains (hereafter referred to as the damaged dataset), and 3) 97 pollen grains from Holocene and Pliocene age (hereafter referred to as the fossil dataset) (**Figure 5.1**). In the fossil dataset, pollen grains also have an altered exine and mineral inclusions (**Figure 5.1**). Using specimens



**Figure 5.1** - Pollen images with automatic z-stack of 128 x 128 pixels for all pollen types of the intact, damaged, and fossil pollen datasets. The name under each image is the final label assigned by the multiple convolutional neural networks (multi-CNNs) with augmentation.

from the reference collection of East African pollen species hosted at CEREGE, Aix-en-Provence, France (**Annexe table 5.1**), eight pollen types were imaged to constitute the training dataset, the intact dataset, and the damaged datasets. These types are *Achyranthes*-type *aspera*, *Aerva*-type *javanica*, *Aerva*-type *lanata*, *Celosia*-type *argentea*, *Cyathula*-type *orthacantha*, *Digera*-type *muricata*, Poaceae, and Cyperaceae. The fossil pollen dataset includes only five of the eight pollen types available in the reference collection due to the limited availability of fossil material. These types are *Achyranthes*-type *aspera*, *Aerva*-type *javanica*, *Digera*-type *muricata*, Cyperaceae, and Poaceae (for sampling locations, depth, and ages see **Annexe table 5.1**). The damaged and fossil datasets were built using samples which were not considered for the training and intact pollen datasets to avoid biases in the estimation of the effectiveness of our approach.

Each pollen grain was imaged under natural light by taking a stack of pictures at 1  $\mu\text{m}$  steps between the top and bottom surfaces of the grain, using an automated custom Leica DMRBE microscope controlled with a software. This step is semi-automated, whereby the image stack movement (z-axis) is automatic but setting the position (x and y axes) is performed manually. Since the stacking step is fixed to 1  $\mu\text{m}$ , the number of fused images is proportional to the size of the pollen. The number of images in the stack ranged from 20 to 40 depending on pollen type. Each stack was split into two halves, and only the half corresponding to the top hemisphere of each pollen grain was kept. The image stack was fused to create a single in-focus image (z-stacked imaging process) with the Helicon Focus 6 software, using the “depth map” method with radius set to 50 pixels and smoothing set to 10 pixels. We chose 1  $\mu\text{m}$  steps empirically, as above this threshold the morphological information was less visible in the fused images and below this value, aberrations appeared into fused images. Each fused image was analyzed using a custom script written for the Fiji image analysis software, to center and crop the pollen into a square image (Schindelin et al., 2012). Finally, all reframed fused images were resized to 128 x 128 pixels.

### 5.2.2. Image classification

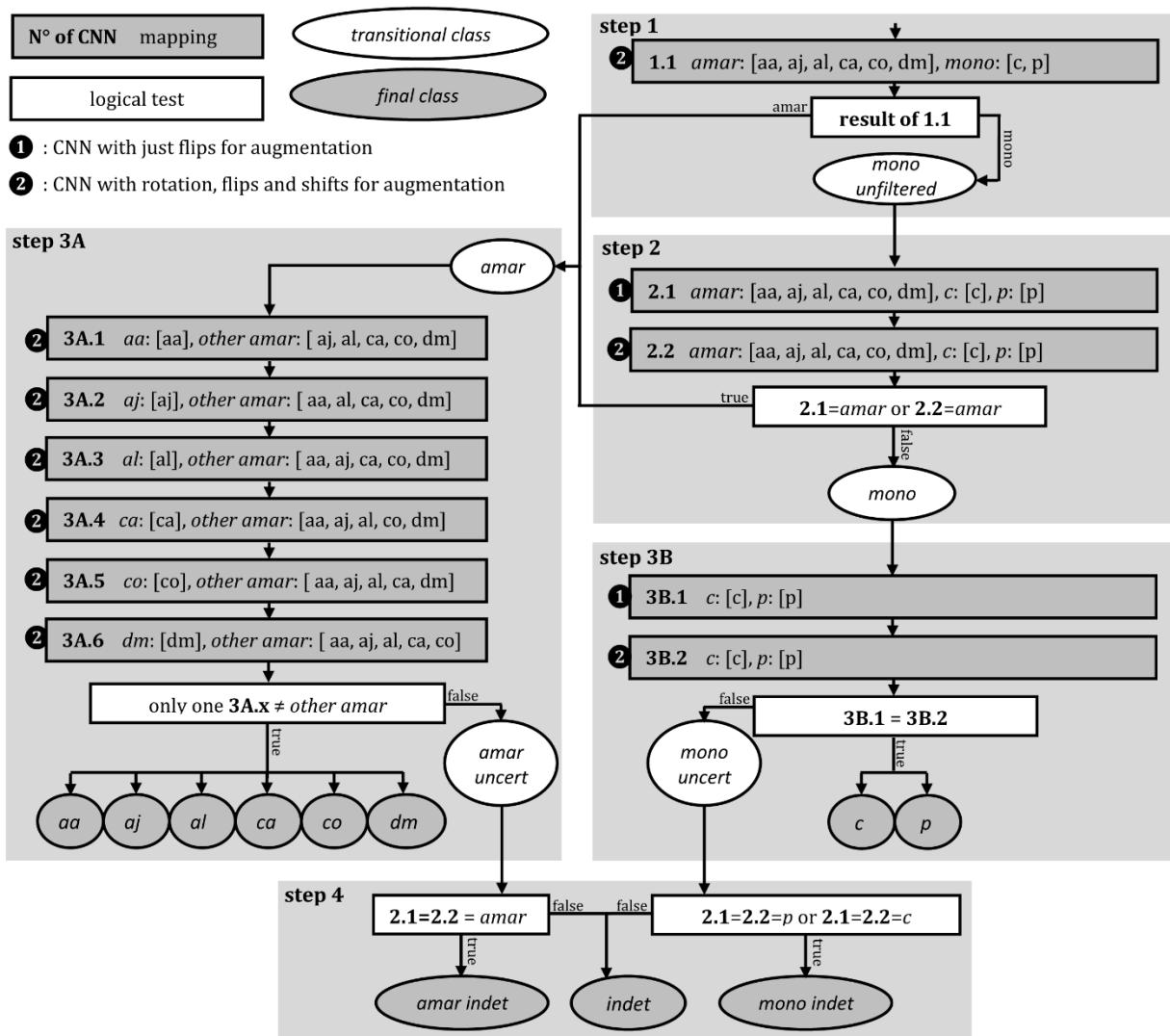
Our system of automatic recognition is developed in the Python language (Pérez et al., 2011) and principally uses the Keras and TensorFlow open-source software libraries (Abadi et al., 2016; Chollet, 2019). It uses convolutional neural networks (CNNs) with a

linear architecture similar to VGGNet (Simonyan and Zisserman, 2014). The CNN training was done with the training pollen dataset and the validation dataset with intact pollen. CNNs training was carried out on batches of 101 pictures (out of a grand total of 1102 pictures) over 500 epochs (the learning algorithm is an iterative process; an epoch is one training iteration on all pictures). To optimize the convergent behavior in loss function, which ensures a more precise adjustment of the weights of our network, we used regressive learning rate as a function of the epoch such as epoch [0; 200] = 0.001, epoch [201; 300] = 0.0006, epoch [301; 400] = 0.00036, and epoch [401; 500] = 0.000216.

We added an augmentation procedure to most of our trainings to generate new images at each epoch. This procedure generates new images by modifying the originals through random rotations (0°–90°), random horizontal and vertical flips, and random width and height shifts (0%–25%). But learning often did not progress when using augmentation because the generation of new images at each epoch made it difficult for the CNN to extract features. To solve this problem, we pre-trained the CNN without augmentation to extract features in its initial layers, then continued training using augmented images. By using this pretrained network for the training with augmentation, the CNN learns to combine these features at the end of the initial layers. This process is commonly called transfer learning (Daood, 2018) (**Annexe figure 5.1**).

The CNNs trained with augmentation are integrated in a decision tree to form a multiple convolutional neural network (multi-CNNs) (**Figure 5.2**). The multi-CNNs is composed of five main steps. Step 1 aims at isolating Amaranthaceae from Monocotyledoneae. Step 2 evaluates if pollen identified as Monocotyledonae in step 1 could be altered Amaranthaceae. Indeed, altered pollen of Dicotyledoneae which have lost their ornamentations and part or all their outer layer (ectexine) may show similarities with Monocotyledonae that only have one layer of exine. To do this in step 2, we used a finer mapping with a separation of Poaceae and Cyperaceae which exhibit different morphologies when intact or slightly altered. Altered Monocotyledonae such as Poaceae and Cyperaceae may easily lose parts of their exine and be deformed or broken, because they are made of a single thin layer. To optimize the chances of recognition of altered pollen, two networks with the same mapping were used but with different augmentation, one focused on shape (CNN 2.1), and the other one on texture (CNN 2.2). In step 3A, several CNNs were used; each one being specialized in the recognition of one pollen type

of Amaranthaceae to maintain CNN high accuracy. The results of all these specialized CNNs are then compiled, the pollen is classified into one Amaranthaceae pollen type or as Amaranthaceae uncertain. In step 3B, the same logic is used as for step 2: two networks with the same mapping are used with different augmentation to separate Poaceae and Cyperaceae. If the results of the two CNNs are incongruous, the pollen is classified as an uncertain Monocotyledoneae. Finally, in step 4, we use the results of the CNNs 2.1 and 2.2 to classify Monocotyledoneae uncertain and Amaranthaceae uncertain in Monocotyledoneae indeterminable, Amaranthaceae indeterminable, or indeterminable pollen.



**Figure 5.2** - Schematic of the classification system based on multiple convolutional neural networks (multi-CNNs) with augmentation. aa: *Achyranthes*-type *aspera*; aj: *Aerva*-type *javanica*; al: *Aerva*-type *lanata*; amar: Amaranthaceae; c: Cyperaceae; ca: *Celosia*-type *argentea*; co: *Cyathula*-type *orthacantha*; dm: *Digera*-type *muricata*; mono: Monocotyledonae; p: Poaceae; uncert: uncertain; indet: indeterminable.

To evaluate the efficiency of augmentation and multi-CNNs, we tested three other approaches: (i) one using multi-CNNs without augmentation in step 1 and 3A, (ii) one simple-CNN without augmentation, and (iii) another simple-CNN with full augmentation of random rotation of 0°–90°, random horizontal and vertical flip, and random width and height shift of 0–25%. The number of palynomorphs analyzed for each pollen type in the intact, damaged, and fossil sets is unbalanced. The number of fossil grains for each type, in particular, is low. As such, we used the average per-class accuracy (APC%) as it is more appropriate for the evaluation of the algorithm performance in this case (Zheng, 2015).

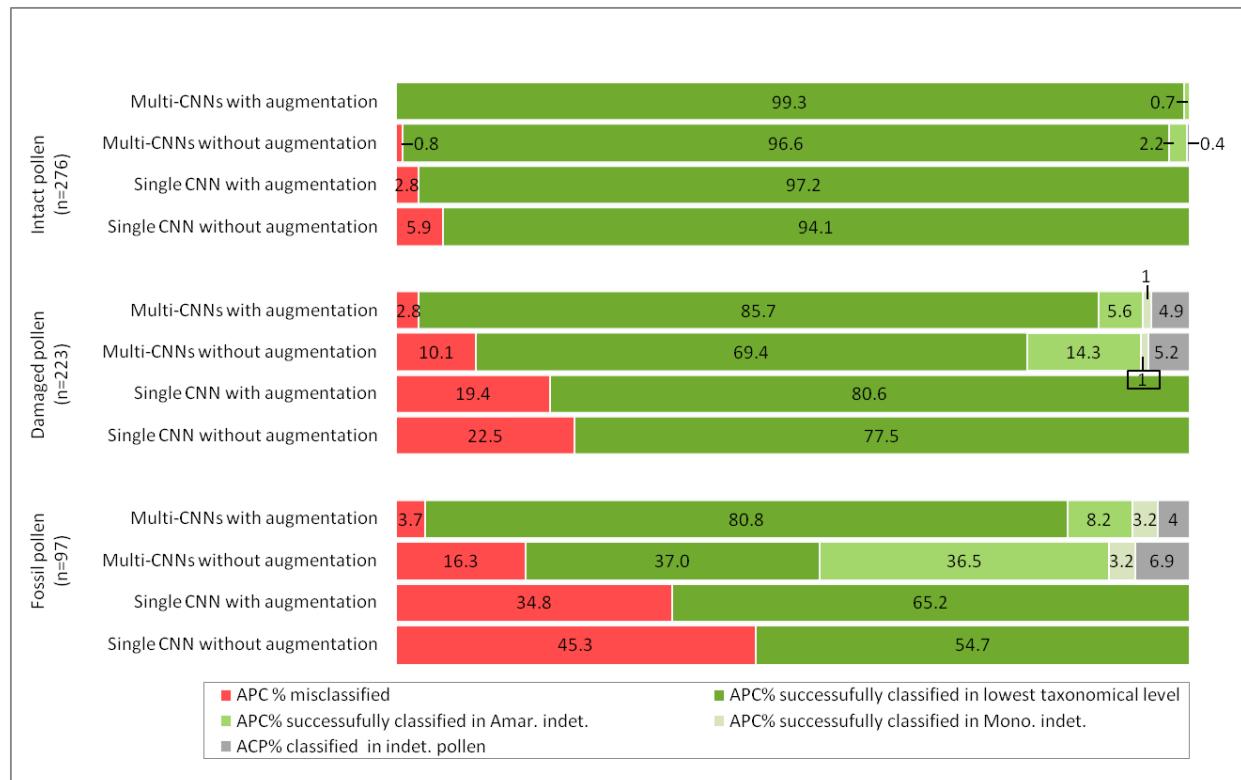
## 5.3. Results

### 5.3.1. Global comparison of used/unused augmentation and single-CNNs/multi-CNNs

The four different classification setups yield misclassification rates of 0%–5.9% for intact pollen, 2.8%–22.5% for damaged pollen, and 3.7%–45.3% for the fossil pollen. The multi-CNNs including augmentation always gives the best results (**Figure 5.3**). Compared with the three other setups, the one using augmentation and multi-CNNs increases the success rate for classification in the lowest taxonomic levels (*Achyranthes-type aspera*, *Aerva-type javanica*, *Aerva-type lanata*, *Celosia-type argentea*, *Cyathula-type orthacantha*, *Digera-type muricata*, Poaceae, and Cyperaceae) and keeps classification rates at 99.3% for intact pollen, 85.7% for damaged pollen, and 80.8% for fossils pollen (**Figure 5.3**).

In detail, augmentation increases the success rate for classification in lowest levels for simple-CNN and multi-CNNs by +3.1% and +2.7% for intact pollen, +3.1% and +16.3% for damaged pollen, and +10.5% and +43.8% for fossil pollen, respectively. Image augmentation also reduces misclassification rates for simple-CNN and multi-CNNs by -3.1% and -0.8% for intact pollen, -3.1% and -7.3% for damaged, and -10.5% and -12.6% for fossil pollen, respectively (**Figure 5.3**). Similarly, the use of multi-CNNs reduces misclassification rates, for the scripts without augmentation (-5.1% for intact pollen, -12.4% for damaged, and -29% for fossil pollen), and for those with augmentation (-2.8% for intact pollen, -16.6% for damaged, and -31.1% for fossil pollen) (**Figure 5.3**).

Multi-CNNs with augmentation also increase the detection of Amaranthaceae and Monocotyledonae misclassification in step 1 and 2. For the damaged dataset, five misclassified pollen grains of Monocotyledonae were directed into step 3A normally reserved to Amaranthaceae (**Annexe figure 5.2.a**). All of these five misclassified pollen grains of Monocotyledonae were identified as Amaranthaceae uncertain in step 3A (**Annexe figure 5.2.b**) then classified as pollen indeterminate in final step 4 (**Figure 5.2**). This is also observed for fossil data in Step 3A and Step 3B (**Annexe figure 5.2.c and .d**).



**Figure 5.3** - Results of our classification system (multi-CNNs with augmentation) as a function of pollen dataset and in comparison with other setups. Pollen types of the lowest taxonomic levels: *Achyranthes*-type *aspera*, *Aerva*-type *javanica*, *Aerva*-type *lanata*, *Celosia*-type *argentea*, *Cyathula*-type *orthacantha* and *Digera*-type *muricata*, and Poaceae and Cyperaceae. indet.: indeterminable. The APC% is the percentages based on average per-class.

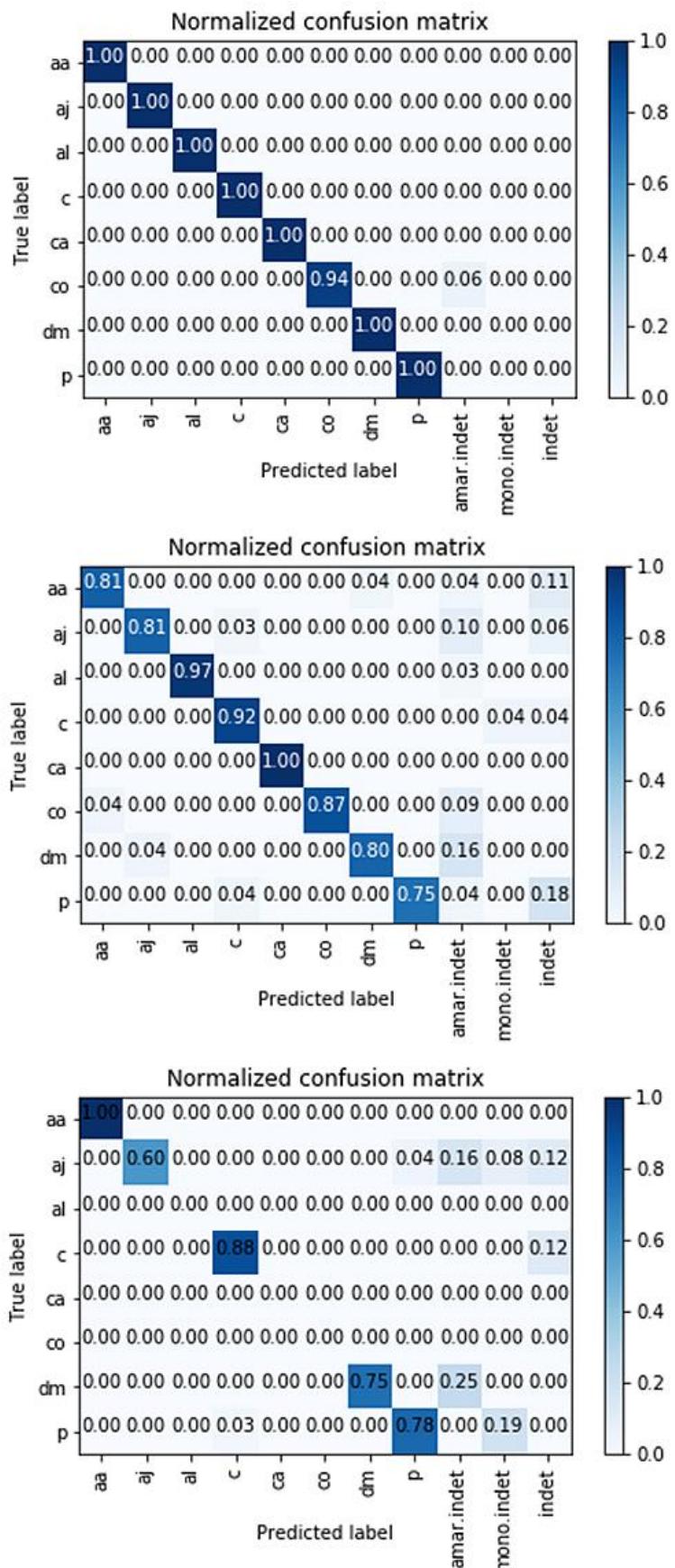
### 5.3.2. Details of efficiency of Multi-CNNs with augmentation

Our Multi-CNNs with augmentation classified the intact pollen grains with misclassification rates of 0%. All intact pollen are correctly classified (**Figure 5.3**). We show that within the Amaranthaceae family, six types with similar morphologies can be distinguished. Over 208 images of intact Amaranthaceae pollen, 99.3% were successfully

classified within different Amaranthaceae types and only 0.7% (two *Cyathula*-type *orthacantha*) could not be clearly identified but could still be defined as Amaranthaceae indeterminable (**Figure 5.3** and **Figure 5.4.a**). For Poaceae and Cyperaceae, all were correctly classified as Poaceae and Cyperaceae despite the heterogeneity of textures and features, and the complex shape of Cyperaceae grains.

The damaged pollen grains, however, show a higher misclassification rate than intact pollen with 2.8% misclassification (**Figure 5.3**). Like for intact pollen, the misclassifications are evenly distributed to different classes (**Figure 5.4.b**). 85.7% of the damaged pollen have been successfully classified to the lowest taxonomic levels (*Achyranthes*-type *aspera*, *Aerva*-type *javanica*, *Aerva*-type *lanata*, *Celosia*-type *argentea*, *Cyathula*-type *orthacantha* and *Digera*-type *muricata*, and Poaceae and Cyperaceae) and 6.6% have been successfully classified to higher taxonomic ranks (5.6% in Amaranthaceae indeterminate, and 1% in Monocotyledonae indeterminate), that is -13.6% and +5.9% than intact pollen, respectively. Of the 167 pollen classified as Amaranthaceae, only 6.6% were classified as Amaranthaceae indeterminate, 6.6% were classified as pollen indeterminate, and none were misclassified as Monocotyledonae indeterminate (**Figure 5.3** and **Figure 5.4.b**).

The fossil grains were classified with 3.7% misclassification, a +0.9% increase compared to damaged pollen (**Figure 5.3**). In this case, misclassifications are also evenly distributed into different classes (**Figure 5.4.c**). 80.8% of total pollen are successfully classified to the lowest taxonomic levels, compared to -4.9% for damaged pollen. We have one increase of pollen classified to higher taxonomic ranks (+4.8%) compared with damaged pollen. We also show a slight decline of pollen classified as pollen indeterminate with a rate of 4% (**Figure 5.3**). The presence of mineral inclusions does not seem to interfere with identification (**Figure 5.1**). Finally, out of the 30 pollen classified as Amaranthaceae, 17% were classified as Amaranthaceae indeterminate.



**Figure 5.4** - Confusion matrix for the intact (a), damaged (b) and fossil (c) pollen dataset for final classes. aa: *Achyranthes*-type *aspera*; aj: *Aerva*-type *javanica*; al: *Aerva*-type *lanata*; amar: Amaranthaceae; c: Cyperaceae; ca: *Celosia argentea*; co: *Cyathula*-type *orthacantha*; dm: *Digera*-type *muricata*; mono: Monocotyledonae; p: Poaceae; indet: indeterminable.

## 5.4. Discussion

Our model has been tested on several datasets with success. The overall computing time required to setup automatic recognition (time required to do all trainings) was about 11h using the Nvidia Titan XGPU for the computing. After the training, each test (inference) takes less than a minute on the same machine, and reproducibility tests show slight variations of <2 APC% (not shown). Our study uses image datasets for the “training dataset”, “intact dataset”, and “damaged dataset” which sizes are in the range of the automatic recognition studies mentioned above, but some may argue that our “fossil dataset” is relatively small. It is, however, a good test for our model because the proportions of Amaranthaceae, Cyperaceae, and Poaceae pollen in the fossil image dataset are close to proportions that are usually observed in modern and fossil pollen assemblages of about 300–400 grains from arid areas of northeastern Africa. The results on the fossil data give a good overview of what can be obtained on real fossil samples. One caveat, not addressed in our study, is that usually, not all pollen types in the fossil/modern sediment or aerial trap are known. As such, the future automatic recognition systems will have to include in the training dataset pollen types that are not related to any of the analyzed/tested types. We are confident, however, that our system should perform well because in our experiment on the fossil pollen dataset, which had none *Aerva*-type *lanata*, *Celosia*-type *argentea*, and *Cyathula*-type *orthacantha*, we reached significant levels, suggesting that including unknown pollen types should not be a major problem in future applications.

Our study shows that a single CNN is not able to manage the recognition of 8 pollen types when the dataset includes damaged or fossil pollen. With a single CNN, misclassification considerably increases with increased grain deterioration. The origin of this problem is linked to the attempt to classify into low taxonomic ranks highly degraded pollen which have lost most of their surface features and ornamentation. This loss of information complicates the recognition by the CNN when the CNN has been trained to recognize eight pollen types. The new approach we designed and which combines the use of augmentation and multi-CNNs, increases the accuracy of classification for the damaged and fossil pollen. This is principally related to the reclassification of highly degraded pollen into higher taxonomic ranks, and the use of CNNs with augmentation and specialized in the recognition of only one pollen type at a time for the lower taxonomic

ranks. The multi-CNNs system is also useful for identifying and self-correcting a significant part of these misclassifications at the high taxonomic level in steps 1 and 2, with the support for the systems set up in steps 3A and 3B. The need for this self-correction is negligible for intact pollen due to the absence of misclassification at the high taxonomic level, but is important for the damaged and fossil pollen which do have misclassification at the high level due to their deteriorated features. Consequently, it is very relevant to use augmentation and a multi-CNNs system for the recognition of fossil or damaged pollen grains.

Our approach based on multi-CNNs with augmentation classifies each class of intact pollen with 0 APC% error. This is as good as in recent studies on intact pollen with a similar number of pollen types (Daood, 2018; Daood et al., 2018; Khanzhina et al., 2018; Sevillano and Aznarte, 2018). For the damaged pollen, we achieved misclassification rate of 2.7 ACP%, which is slightly higher than that of intact pollen (**Figure 5.1**). To our knowledge, this is the first time that such good accuracy on highly damaged pollen grains has been reached (Han and Xie, 2018; Holt and Bennett, 2014; Khanzhina et al., 2018). Based on all human-made pollen counts of actual soil and mud samples in the African Pollen Database available for Ethiopia (115 samples), we find that indeterminable (i.e. damaged) pollen represent on average about 8% of total pollen count, and that indeterminable Amaranthaceae are most common: they represent about 35% out of total Amaranthaceae (“African Pollen Database,” 2019). Our percent of indeterminable pollen out of the total pollen (2.7 APC%) and of indeterminable Amaranthaceae out of the total Amaranthaceae (5.6 APC%) are thus better than the human counts, for modern damaged pollen in the Ethiopian dataset (Abadi et al., 2016; Bonnefille, 1972; Buchet, 1982).

As mentioned before, Punyasena et al. (2012) and Kong et al. (2016) obtained good results for fossil pollen of *Picea*, but the grains in their study came from a peat; altered pollen in peat cores are far from resembling altered fossil pollen grains from lacustrine Pliocene samples. Our approach is the first to present successful results (only 4.1% misclassification) for fossil pollen with important deterioration, typical for arid areas. In the human-made Pliocene pollen dataset of about 25 samples obtained for the Hadar formation by Bonnefille’s team (Bonnefille et al., 1987) indeterminable pollen may account for up to about 30% of total pollen counts (10% in average). Of these 25 samples, 19 have Amaranthaceae of which about 80% on average were classified as

“Amaranthaceae indeterminable” (Bonnefille et al., 1987). The percentage of fossil pollen indeterminate that we obtained with our automated system is of 4%, i.e. well within the average of Bonnefille et al. (1987), and the taxonomic precision is better as only 17% of Amaranthaceae were classified as indeterminable. Our method, therefore, has the potential to be applied to both damaged, fossil, and modern pollen grains from arid areas, and may even improve pollen taxa identifications within the Amaranthaceae.

Using our multi-CNNs with augmentation, we were able to classify both morphologically similar pollen (pollen types of Amaranthaceae) at the species or genus level, and morphologically different pollen (Amaranthaceae versus Poaceae versus Cyperaceae), without having significant misclassification for intact, damaged, and fossil pollen (**Figure 5.4**). For the Monocotyledonae, we can observe an increase in fossil pollen classified as indeterminable Monocotyledonae (+6.8%) in comparison to the damaged pollen, but we do not have significant increases for Amaranthaceae (**Figure 5.3**). The fossil grains of Monocotyledonae have more mineral inclusions than Amaranthaceae, but we discard the hypothesis that mineral inclusions in pollen is the origin of these increases because fossil pollen of Monocotyledonae bearing mineral inclusions are correctly identified (**Figure 5.1**). We see on the images that fossil pollen of Monocotyledonae is often “stuck” in residual sediments (**Figure 5.1**). This is very similar to images of “contaminated grains” of Han and Xie (2018), who suggested that this hinders identification. Our study shows that our system overcomes the sediment contamination and is able to correctly identify fossil pollen of Amaranthaceae even when sediment is attached to the grains. The deformation of pollen grains is another hindrance to identification (Han and Xie, 2018; Holt et al., 2011; Khanzhina et al., 2018) but our fossil grains of Monocotyledonae are no more deformed than damaged grains of Monocotyledonae or fossil grains of Amaranthaceae (**Figure 5.1**). We posit that, within the biodiversity range we investigated, this issue stems from the structural weakness of Monocotyledonae (in the pollen studied) as this family shows a thinner and weaker exine than Amaranthaceae. As for human eyes, automated pollen recognition requires that the exine texture must be relatively well preserved. This agrees with the idea of Li et al. (2004) that the “surface texture is frequently characteristic, even when grains are damaged or fragmented” for the pollen.

Our multi-CNNs with augmentation is operational. At CEREGE, it is used occasionally to assist with identification of fossil Amaranthaceae when identification to type is not straightforward. The main bottleneck to an extended use of our recognition system is the acquisition time, which is manual and therefore, time-consuming, preventing an easy addition of new pollen types. Adding an automated acquisition step would therefore provide a complete automatic pollen counting system, saving a lot of time for palynologists of arid areas and possibly elsewhere. With such a system, it would probably be easy to acquire images of large pollen reference collections within a few months. The prospect of a fully automated pollen counting system for paleoclimatic and paleoenvironmental reconstructions, similar to the operating systems developed for other microfossils like coccoliths or foraminifera (Beaufort and Dollfus, 2004; Marchant et al., In Preparation), can be foreseen in the very near future. These automated systems will pave the way to high-throughput palynology studies, including morphometric studies.

## 5.5. Conclusions

In this study, we proposed a robust system for classification of very damaged and badly preserved fossil grains. We have developed a classification topology that has flexibility in the choice of taxonomic levels used for classification. It can, like a palynologist, assign higher taxonomic ranks if the characteristics of the image are not sufficient for robust identification at a lower taxonomic level. Moreover, if images cannot be properly classified, it can declare them as indeterminable. Our system has the advantage of using simple LM images and does not require the implementation of complex and expensive systems such as confocal fluorescence microscopes. With the damaged pollen, we have obtained results similar to or better than those obtained from human pollen counting performed by CEREGE palynologists on modern soils and Pliocene sediments of the Awash Valley in Ethiopia.

The system developed here does not yet allow a complete count of fossil pollen. It remains limited to a few families and to a very limited number of taxa in comparison to overall pollen diversity. Evaluating and improving how the system reacts to pollen types or other grains not included in the training set is also important because usually not all

pollen types in the sediment core are known. To handle this, one solution may be the addition of an “unknown” category. Our results, based on a limited dataset but representing a wide range of biases affecting pollen grains from their production to their fossilization, allow the possible development of automatic pollen recognition for routine palynological studies of samples coming from multiple depositional environments and with different degrees of pollen alteration. It would also be interesting to see if it is possible to increase the accuracy of our system with the use of a much larger number of training images. It is also important to specify that the system presented here cannot be directly extended to other pollen families as it relies more on the morphology of the pollen types than on the taxonomic relationship, although the two are related. It is also important to have a good knowledge of how the morphology of each type of pollen changes with degradation or fossilization in order to build relevant mappings and logical steps in the classification system to cover all stages of pollen grain degradation. The expertise of paleo-palynology is therefore critical to the extension of this system.

The next step of this work will be to extend the system to automatically take pictures of pollen grains from microscope slides. The automatic acquisition will be inspired by the processing software developed for coccoliths at CEREGE (Beaufort and Dollfus, 2004) and modified by Martin Tetard and Yves Gally for Radiolaria tests (Martin Tetard, personal communication). It can be summarized in 6 steps: (1) color pollen grains with safranin for easier segmentation, (2) take grayscale images of microscope fields of view (FOVs) by incrementally stepping through the slide ( $1 \mu\text{m}$  step), (3) segment each pollen grains in each stacked FOV (saving their x and y coordinates on the original unstacked FOV), (4) generate sub-stacks for each previously recorded pollen grain from each original unstacked FOV, (5) remove the unfocused pollen images from each pollen sub-stack using a gray level threshold, and (6) apply the method in the second paragraph of part 2.1 for finalizing image processing prior to automated recognition. With this automatic acquisition, it will be possible to fully scan reference pollen collections and thus quickly build a large database of pollen images to develop an automatic pollen counting system (acquisition and recognition) for modern and fossil pollen from eastern Africa, for example.

## **5.6. Computer code availability**

All files and folders presented here are downloadable at <https://github.com/Beniofh/Multi-CNN>. We used the script Multi-CNN\_test\_V3.5.2. py (40.9 Ko) to recognize pollen types using multi-CNNs with augmentation. Multi-CNNs with augmentation is compared to single- CNN with augmentation (script CNN\_test\_v2.1.0. py, 16.5 Ko), single- CNN without augmentation (CNN\_test\_v2.1.0 (wihout\_augm). py, 16.5 Ko), and multi-CNNs without augmentation (Multi-CNN\_test\_V3.5.2 (wihout\_augm). py, 40.9 Ko). In the CNN\_solo folder, we have for each CNN model parameters, a folder with file of CNN model parameter used in .ckpt, histograms of accuracy, loss and learning rate during the training in .bmp and table of accuracy, loss and learning rate during the training in .csv formats. The folders “intact”, “damaged” and “fossil” contain respectively the images for “training dataset”/“intact pollen dataset”, “damaged dataset”, and “fossil dataset”. The User\_manual.pdf explains how to use these scripts to reproduce our results. These scripts have been developed by Benjamin BOUREL (benjaminbourel1@gmail.com; +33 6 63 49 85 24) with the help of Ross MARCHANT. These scripts were first available the 13th October 2019. They were created in the open-source Anaconda 3 v.5.2.0 (<https://www.anaconda.com>) with Spyder v.3.3.6 in python v.3.6 (Pérez et al., 2011). We also used NumPy (Oliphant, 2015), Scikit-image (van der Walt et al., 2014), Scikit-learn (Pedregosa et al., 2011), and Matplotlib (Hunter, 2007) packages. The details of the python libraries used (name and version) are indicated in User\_manual.pdf. All software and python libraries used here are open sources. For optimal use, please use the indicated versions.

## **5.7. Acknowledgments**

We thank J-C. Mazur of the CEREGE’s micropaleontology platform for help during pollen sample preparations, and F. Chalié for the automated microscope. We also thank G. Gassier for help with the server, and M. Djamali for his pollen slides from Urima Lake (Iran). This work was supported by Aix-Marseille University (doctoral fellowship to BB), the research grant ANR PRCE FIRST ANR-15-CE04-0006-01. The research leading to these results has received funding from the People Program (Marie Curie Actions) of the European Union’s Seventh Framework Program (FP7/2007–2013) under REA grant

agreement n. PCOFUND-GA-2013-609102, through the PRESTIGE program coordinated by Campus France.

## 6. Synthèse

Ce travail a permis d'apprécier l'efficacité du proxy multifactoriel que sont les pollens pour caractériser les environnements modernes d'Afrique de l'Est. Nous avons été capables de caractériser des végétations en dessous de l'échelle régionale et de séparer le signal pollinique local du signal régional grâce au développement d'une nouvelle approche de calcul de scores polliniques basée sur des analyses statistiques (*T-values* et *hierarchical clustering*) et de l'apprentissage automatique (*k-means partitioning*). Cela nous a permis de proposer pour la première fois une approche analytique capable d'identifier les végétations azonales (qui ne sont pas en équilibre avec le climat) et locales (restreintes dans le paysage par les conditions édaphiques). Ces végétations particulièrement importantes dans les zones arides du rift est africain sont liées à des résurgences d'eaux souterraines ou à des rivières. Nous avons pu constater sans grande surprise l'efficacité de cette approche également pour inférer à partir des spectres polliniques les paramètres classiques du climat, à savoir températures et précipitations. Mais, à notre surprise, cette approche a aussi permis de renseigner des paramètres plus subtils comme la saisonnalité des températures, la saisonnalité des précipitations, les régimes de précipitations ou le pourcentage du couvert arboré. Au niveau des végétations riveraines, il a même été possible d'inférer la morphologie (méandrique ou rectiligne) et la nature (saisonnière ou pérenne) des rivières. Ce mémoire met en avant le potentiel inégalé des pollens pour la reconstruction des paléoenvironnements en Afrique de l'Est comparé aux autres proxys classiquement utilisés que sont les fossiles de macrofaunes, le  $\delta^{13}\text{C}$ , les phytolithes ou les composés organiques (n-alcanes) provenant des cires foliaires qui, bien souvent, ne permettent d'inférer qu'un seul paramètre (eg., la densité du couvert arboré).

L'approche analytique développée ici a été appliquée au registre fossile, mais elle reste encore à améliorer. En effet, à ce jour seuls 67 échantillons de surface permettent de documenter les végétations azonales alors que celles-ci sont réparties sur plus de 2000 km le long du rift (juste pour les rivières Awash et Omo) et correspondent certainement à des formations édaphiques avec des compositions floristiques particulières. Echantillonner davantage de végétations de résurgences d'eaux comme les résurgences

saisonnieres permettrait notamment d'affiner les reconstructions pour les résurgences qui ne comportent actuellement que des résurgences pérennes. L'échantillonnage doit aussi être poursuivi au niveau des vases modernes (de surface) de rivières et de lacs, car ces échantillons sont sous-représentés par rapport aux échantillons de sols modernes. Les sédiments des sites à Hominini plio-pléistocènes du rift est africain sur lesquels cette approche doit être appliquée sont majoritairement des sédiments de rivières, de lacs et de deltas. Il n'est donc pas raisonnable de continuer à interpréter des spectres polliniques fossiles de vases à l'aide de calibrations majoritairement faites à partir de spectres polliniques de sols modernes. Des échantillons provenant de contextes de dépôts actuels similaires aux fossiles méritent d'être plus largement étudiés.

Appliquée aux données fossiles, notre nouvelle approche de calcul de scores polliniques a permis d'apporter une dimension nouvelle aux interprétations des données polliniques fossiles du Pliocène de l'Afar, en permettant de caractériser plus clairement les habitats potentiellement associés à *Australopithecus afarensis* dans la basse vallée de l'Awash (de 3,42 Ma à 2,96 Ma). Nos résultats montrent que cet ancêtre de la lignée humaine avait accès, localement, au sein même de biomes ouverts et arides/semi-arides à des micro-habitats tels que les forêts édaphiques et ripariennes. Par analogie avec l'actuel, ces microenvironnements ou micro-habitats distincts et variés sont largement plus boisés que la végétation à l'échelle du paysage, puisque inféodés à des résurgences d'eaux profondes et des rivières, mais leur extension spatiale est limitée aux zones édaphiques humides. Non seulement ces habitats sont plus boisés mais ils sont aussi plus stables que la végétation à l'échelle régionale qui est soumise aux contraintes climatiques. Ainsi, nos résultats contribuent à remettent en cause deux paradigmes : celui qui positionne l'évolution de la bipédie dans un contexte d'ouverture croissante des milieux (expansion des savanes) (Cerling et al., 2011; Lamarck, 1809), et celui qui lie obligatoirement ou exclusivement l'évolution morphologique aux changements environnementaux (Maslin and Trauth, 2009).

Cette remise en cause d'une écologie ubiquiste par une écologie plutôt forestière n'est pas en opposition avec les analyses des restes postcrâniens d'*Australopithecus afarensis* (Ruff et al., 2016). Elle est aussi cohérente avec les études écologiques sur les primates modernes. En effet, les primates actuels vivant dans des végétations régionales ouvertes et arides/semi-arides ont leur domaine vital (*home range*) au niveau des forêts

riveraines qui traversent ces végétations régionales. Il est alors raisonnable de penser que cela aurait aussi pu être le cas pour *Australopithecus afarensis*. Tout ceci va permettre de débattre des hypothèses liées aux savanes et de poursuivre les discussions sur la nature plésiomorphe ou apomorphe des caractères arboricoles visibles chez *Australopithecus afarensis*. Il reste encore à effectuer un travail similaire sur sites où sont référencés d'autres genres d'Hominini contemporains d'*Australopithecus afarensis* dans la basse vallée de l'Awash. Il faut aussi poursuive la recherche d'échantillons polliniques à Hadar pour la période correspondant à la fin de l'occurrence d'*Australopithecus afarensis* dans cette zone. Ceci pour voir s'il y a des changements à l'échelle locale, et notamment s'il est possible d'observer une disparition ou une réduction du signal pollinique des forêts et environnements ripariens. Cela permettrait de renforcer l'idée de leur affiliation à ce type d'environnement.

Il sera aussi nécessaire de poursuivre les investigations au niveau des résurgences d'eaux associées aux sites de la basse vallée de l'Awash. Nous avons analysé un système de résurgence moderne potentiellement analogue à celui qui était présent dans les niveaux à Hominini d'Olduvai. Cette analyse réalisée sur une courte carotte (43 cm) a permis de bien comprendre, à l'échelle du millénaire, les dynamiques de la végétation sur le site et par extension celle de la résurgence. Elle a montré que le type de système de résurgence présent à Olduvai pouvait maintenir une activité continue malgré les variations climatiques, permettant ainsi de supporter le développement démographique et culturel des populations d'Hominini présentes. Cependant, il existe plusieurs types de systèmes de résurgences et ceux d'Hadar et Dikika semblent différents de ceux d'Olduvai. Il faut donc répéter ce type d'étude sur d'autres résurgences dans le rift qui ont potentiellement un fonctionnement différent pour voir si la dynamique de leur végétation est similaire à celle observée dans ce manuscrit. Et ce n'est pas le cas, il faudra voir pourquoi on observe des différences (systèmes hydrologiques différents, climats différents, sols différents, faunes différentes, etc.).

L'une des plus grandes difficultés rencontrées durant cette thèse est l'identification et le dénombrement des taxons polliniques au microscope optique. Cette thèse a demandé d'apprendre à reconnaître et interpréter un peu plus de 400 types polliniques répartis dans 140 familles végétales. Le temps de comptage a également été un obstacle important et a limité l'acquisition des données. Les échantillons polliniques

comptés durant cette thèse ont requis entre 2 et 5 jours en moyenne selon la qualité de la préservation des grains de pollens et leur abondance dans le sédiment. Bien que très informatifs, les pollens sont cependant un proxy très peu compétitif par rapport à d'autres, moins informatifs mais dont l'acquisition a pu être automatisée (ex. le  $\delta^{13}\text{C}$ ). Ainsi, nous avons dû nous résoudre à ne pas compter au moins 13% des échantillons des carottes HSPDP-NAW et NAO car la qualité et l'abondance des grains de pollens auraient demandé plusieurs semaines de travail par échantillon. Si nous avions pu les exploiter, nous aurions eu des comptages pour presque 20% des échantillons de NAW14-1A et NAO14-1B au lieu de 5% seulement.

Dans l'objectif de réduire le temps d'acquisition des données polliniques et d'optimiser l'étude des échantillons sub-stériles, il nous paraît important de développer l'automatisation de la reconnaissance des pollens et des comptages. Cela se fait déjà pour d'autres microfossiles comme les coccolites (Beaufort and Dollfus, 2004), les foraminifères (de Garidel-Thoron et al., 2020) et les radiolaires (Tetard et al., 2020). La preuve de concept que nous avons présentée ici nous encourage à poursuivre dans cette voie. En priorité, il faudrait mettre en place un système totalement automatique pour l'acquisition des images, qui est pour l'instant limité à la segmentation des images et à leur empilement de mises au point (*focus stacking*). Des tests préliminaires ont montré qu'il était possible d'adapter le système d'acquisition automatique du SYstème de Reconnaissance Automatique des COccolithes (SYRACO) développé au CEREGE. Le système de réseau de neurones artificiels développé ici devra également être amélioré notamment en intégrant des réseaux de neurones récurrents (*recurrent neural network*: RNN) dans le système. Ces RNN permettraient de proposer des cubes de données plutôt que des images en 2D. Notre système d'arbre décisionnel est aussi un peu limité, il faudrait qu'il soit assisté par un réseau de type réseaux de neurones à impulsions (*spiking neural network*: SNN) qui serait très bien adapté à cette tache car il limite fortement les faux positifs. L'utilisation d'algorithmes évolutionnaires (*evolutionary algorithms*) sur les paramètres d'augmentation de l'apprentissage des réseaux de neurones serait aussi un atout important.

## Bibliographie

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv160304467 Cs.
- Abram, N.J., Wright, N.M., Ellis, B., Dixon, B.C., Wurtzel, J.B., England, M.H., Ummenhofer, C.C., Philibosian, B., Cahyarini, S.Y., Yu, T.-L., 2020. Coupling of Indo-Pacific climate variability over the last millennium. *Nature* 579, 385–392.
- African Plant Database [WWW Document], 2019. URL [ville-ge.ch/musinfo/bd/cjb/africa/](http://ville-ge.ch/musinfo/bd/cjb/africa/) (accessed 3.25.19).
- African Pollen Database [WWW Document], 2019. . Afr. Pollen Database. URL <http://apd.sedoo.fr/accueil.htm> (accessed 5.6.19).
- Agnew, A.D.Q., Agnew, S., 1994. Upland Kenya Wild Flowers: A Flora of the Ferns and Herbaceous Flowering Plants of Upland Kenya. East Africa Natural History Society, Nairobi.
- Albert, R.M., Bamford, M.K., Esteban, I., 2015. Reconstruction of ancient palm vegetation landscapes using a phytolith approach. *Quat. Int., Eastern African Quaternary Research Association, 4th conference* 369, 51–66. <https://doi.org/10.1016/j.quaint.2014.06.067>
- Alemseged, Z., Spoor, F., Kimbel, W.H., Bobe, R., Geraads, D., Reed, D., Wynn, J.G., 2006. A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature* 443, 296–301. <https://doi.org/10.1038/nature05047>
- Arráiz, H., Barboni, D., Uribelarrea, D., Mabulla, A., Baquedano, E., Domínguez-Rodrigo, M., 2017. Paleovegetation changes accompanying the evolution of a riverine system at the BK paleoanthropological site (Upper Bed II, Olduvai Gorge, Tanzania). *Palaeogeogr. Palaeoclimatol. Palaeoecol., Paleoecological reconstructions of the Bed I and Bed II lacustrine basins of Olduvai Gorge (Tanzania) and insights into early human behavior* 488, 84–92. <https://doi.org/10.1016/j.palaeo.2017.05.010>
- Ashley, G.M., Barboni, D., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Diez-Martin, F., Barba, R., Baquedano, E., 2010a. Paleoenvironmental and paleoecological reconstruction of a freshwater oasis in savannah grassland at FLK North, Olduvai Gorge, Tanzania. *Quat. Res.* 74, 333–343. <https://doi.org/10.1016/j.yqres.2010.08.006>
- Ashley, G.M., Barboni, D., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Diez-Martin, F., Barba, R., Baquedano, E., 2010b. A spring and wooded habitat at FLK Zinj and their relevance to origins of human behavior. *Quat. Res., Paleoecology and Hominin Behavior during Bed I at Olduvai Gorge (Tanzania)* 74, 304–314. <https://doi.org/10.1016/j.yqres.2010.07.015>
- Ashley, G.M., Mworia, J., Muasya, A., Owen, R., Driese, S., Hover, V., Renaut, R., Goman, M., Mathai, S., Blatt, S., 2004. Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Loboi Swamp, Kenya, East Africa. *Sedimentology* 51, 1301–1321. <https://doi.org/10.1111/j.1365-3091.2004.00671.x>

- Ashley, G.M., Tactikos, J.C., Owen, R.B., 2009. Hominin use of springs and wetlands: Paleoclimate and archaeological records from Olduvai Gorge (~1.79–1.74 Ma). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 272, 1–16. <https://doi.org/10.1016/j.palaeo.2008.10.016>
- Ayenew, T., Demlie, M., Wohlnich, S., 2008. Hydrogeological framework and occurrence of groundwater in the Ethiopian aquifers. *J. Afr. Earth Sci.* 52, 97–113. <https://doi.org/10.1016/j.jafrearsci.2008.06.006>
- Bamford, M.K., 2012. Fossil sedges, macroplants, and roots from Olduvai Gorge, Tanzania. *J. Hum. Evol.*, Five Decades after *Zinjanthropus* and *Homo habilis*: Landscape Paleoanthropology of Plio-Pleistocene Olduvai Gorge, Tanzania 63, 351–363. <https://doi.org/10.1016/j.jhevol.2011.07.001>
- Barboni, D., 2014. Vegetation of Northern Tanzania during the Plio-Pleistocene: A synthesis of the paleobotanical evidences from Laetoli, Olduvai, and Peninj hominin sites. *Quat. Int.*, The Evolution of Hominin Behavior during the Oldowan-Acheulian Transition: Recent Evidence from Olduvai Gorge and Peninj (Tanzania) 322–323, 264–276. <https://doi.org/10.1016/j.quaint.2014.01.016>
- Barboni, D., Ashley, G.M., Bourel, B., Arráiz, H., Mazur, J.-C., 2019. Springs, palm groves, and the record of early hominins in Africa. *Rev. Palaeobot. Palynol.* 266, 23–41. <https://doi.org/10.1016/j.revpalbo.2019.03.004>
- Barboni, D., Ashley, G.M., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Baquedano, E., 2010. Phytoliths infer locally dense and heterogeneous paleovegetation at FLK North and surrounding localities during upper Bed I time, Olduvai Gorge, Tanzania. *Quat. Res.* 74, 344–354. <https://doi.org/10.1016/j.yqres.2010.09.005>
- Barboni, D., Bonnefille, R., 2001. Precipitation signal in pollen rain from tropical forests, South India. *Rev. Palaeobot. Palynol.* 114, 239–258. [https://doi.org/10.1016/S0034-6667\(01\)00057-4](https://doi.org/10.1016/S0034-6667(01)00057-4)
- Beaufort, L., Dollfus, D., 2004. Automatic recognition of coccoliths by dynamical neural networks. *Mar. Micropaleontol.* 51, 57–73. <https://doi.org/10.1016/j.marmicro.2003.09.003>
- Behrensmeyer, A.K., Reed, K.E., 2013. Reconstructing the Habitats of *Australopithecus*: Paleoenvironments, Site Taphonomy, and Faunas, in: Reed, K.E., Fleagle, J.G., Leakey, R.E. (Eds.), *The Paleobiology of Australopithecus*, Vertebrate Paleobiology and Paleoanthropology. Springer Netherlands, Dordrecht, pp. 41–60. [https://doi.org/10.1007/978-94-007-5919-0\\_4](https://doi.org/10.1007/978-94-007-5919-0_4)
- Bobe, R., Manthi, F.K., Ward, C.V., Plavcan, J.M., Carvalho, S., 2020. The ecology of *Australopithecus anamensis* in the early Pliocene of Kanapoi, Kenya. *J. Hum. Evol.*, Kanapoi: The Paleobiology of a Pliocene Site in the Turkana Basin, Kenya 140, 102717. <https://doi.org/10.1016/j.jhevol.2019.102717>
- Bonnefille, R., 2018. Sur les pas de Lucy - Expédition en Ethiopie, Odile Jacob. ed. Paris, France.
- Bonnefille, R., 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Glob. Planet. Change* 72, 390–411. <https://doi.org/10.1016/j.gloplacha.2010.01.015>
- Bonnefille, R., 1984. Palynological research at Olduvai Gorge. *Natl. Geogr. Soc. Res. Rep.* 17, 227–243.
- Bonnefille, R., 1979. Méthode palynologique et reconstitutions paleoclimatiques au Cenozoïque dans le Rift est-africain. *Bull. Société Géologique Fr. S7-XXI*, 331–342. <https://doi.org/10.2113/gssgbull.S7-XXI.3.331>

- Bonnefille, R., 1972. Associations polliniques actuelles et quaternaires en Ethiopie (vallées de l'Awash et de l'Omo) (Associations polliniques actuelles et quaternaires en Ethiopie). Paris VI, Paris, France.
- Bonnefille, R., Buchet, G., 1986. Forest history in Ethiopia, a pollen diagram from Wenchi. SINET Ethio J Sci 9, 169–178.
- Bonnefille, R., Buchet, G., Friis, I.B., Kelbessa, E., Mohammed, M.U., 1993. Modern pollen rain on an altitudinal range of forests and woodlands in South West Ethiopia. Opera Bot. 121, 71–84.
- Bonnefille, R., Chalié, F., Guiot, J., Vincens, A., 1992. Quantitative estimates of full glacial temperatures in equatorial Africa from palynological data. Clim. Dyn. 6, 251–257.
- Bonnefille, R., Potts, R., Chalié, F., Jolly, D., Peyron, O., 2004. High-resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. Proc. Natl. Acad. Sci. 101, 12125–12129. <https://doi.org/10.1073/pnas.0401709101>
- Bonnefille, R., Riollet, G., 1988. Palynologie des sediments holocènes de sites archéologiques du Qatar. Rech. Sur Civilis. Préhistoire À Qatar Paris 137–146.
- Bonnefille, R., Riollet, G., 1987. Palynological spectra from the Upper Laetolil Beds, in: The Pliocene Site of Laetoli, Northern Tanzania. Leakey, M.D., Harris, J.M. (Eds.), Oxford, pp. 52–61.
- Bonnefille, R., Riollet, G., Chalié, F., Jolly, D., Peyron, O., 2004. High-resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. Proc. Natl. Acad. Sci. 101, 12125–12129.
- Bonnefille, Vincens, A., Buchet, 1987. Palynology, stratigraphy and palaeoenvironment of a pliocene hominid site (2.9-3.3 M.Y.) at Hadar, Ethiopia. Palaeogeogr. Palaeoclimatol. Palaeoecol. 60, 249–281. [https://doi.org/10.1016/0031-0182\(87\)90035-6](https://doi.org/10.1016/0031-0182(87)90035-6)
- Borcard, D., Gillet, F., Legendre, P., 2018. Numerical Ecology with R. Springer.
- Bourel, B., Marchant, R., de Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., Beaufort, L., 2020. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. Comput. Geosci. 140, 104498. <https://doi.org/10.1016/j.cageo.2020.104498>
- Brown, E.T., Johnson, T.C., 2005. Coherence between tropical East African and South American records of the little ice age. Geochem. Geophys. Geosystems 6.
- Buchet, G., 1982. Transport des pollens dans les fleuves Omo et Awash (Ethiopie): étude des vases actuelles. École pratique des hautes études, Montpellier, France.
- Campisano, C.J., 2019. Geological collaboration at paleoanthropological sites in Ethiopia's lower Awash Valley. Evol. Anthropol. Issues News Rev. 28, 228–232. <https://doi.org/10.1002/evan.21800>
- Campisano, C.J., 2012. Geological summary of the Busidima Formation (Plio-Pleistocene) at the Hadar paleoanthropological site, Afar Depression, Ethiopia. J. Hum. Evol. 62, 338–352. <https://doi.org/10.1016/j.jhevol.2011.05.002>
- Campisano, C.J., 2007. Tephrostratigraphy and hominin paleoenvironments of the Hadar Formation, Afar Depression, Ethiopia. Rutgers University - Graduate School-New Brunswick. <https://doi.org/10.7282/T3NS0V99>
- Campisano, C.J., Cohen, A.S., Arrowsmith, J.R., Asrat, A., Behrensmeyer, A.K., Brown, E.T., Deino, A.L., Deocampo, D.M., Feibel, C.S., Kingston, J.D., Lamb, H.F., Lowenstein, T., Noren, A., Olago, D., Owen, R.B., Pelletier, J.D., Potts, R., Reed, K.E., Renaut, R.W., Russell, J.M., Russell, J.L., Schäbitz, F., Trauth, M.H., Wynn, J.G., 2017. The Hominin Sites and Paleolakes Drilling Project: High-resolution paleoclimate records from the East African Rift System and their implications for understanding the environmental context of hominin evolution. PaleoAnthropology.

- Campisano, C.J., Feibel, C.S., 2008. Depositional environments and stratigraphic summary of the Pliocene Hadar Formation at Hadar, Afar Depression, Ethiopia, in: The Geology of Early Humans in the Horn of Africa. Geological Society of America, pp. 179–201.
- Carr, C.J., 1998. Patterns of vegetation along the Omo River in southwest Ethiopia. *Plant Ecol.* 135, 135–163. <https://doi.org/10.1023/A:1009704427916>
- Caylor, K.K., Shugart, H.H., 2006. Pattern and Process in Savanna ecosystems, in: D'Odorico, P., Porporato, A. (Eds.), *Dryland Ecohydrology*. Springer Netherlands, Dordrecht, pp. 259–281. [https://doi.org/10.1007/1-4020-4260-4\\_15](https://doi.org/10.1007/1-4020-4260-4_15)
- Cerling, T.E., Hay, R.L., 1986. An isotopic study of paleosol carbonates from Olduvai Gorge. *Quat. Res.* 25, 63–78. [https://doi.org/10.1016/0033-5894\(86\)90044-X](https://doi.org/10.1016/0033-5894(86)90044-X)
- Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., Mace, W., Macharia, A.N., Quade, J., Remien, C.H., 2011. Woody cover and hominin environments in the past 6 million years. *Nature* 476, 51–56. <https://doi.org/10.1038/nature10306>
- Chollet, F., 2019. Keras [WWW Document]. URL <https://github.com/fchollet/keras>
- Chudyk, C., Castaneda, H., Léger, R., Yahiaoui, I., Boochs, F., 2015. Development of an Automatic Pollen Classification System Using Shape, Texture and Aperture Features, in: Proceedings of the LWA 2015 Workshops: KDML FGWM IR and FGDB Trier. Presented at the Proceedings of the LWA 2015 Workshops: KDML FGWM IR and FGDB Trier, Trier, Germany, pp. 65–74.
- Cohen, A., Campisano, C., Arrowsmith, R., Asrat, A., Behrensmeyer, A.K., Deino, A., Feibel, C., Hill, A., Johnson, R., Kingston, J., Lamb, H., Lowenstein, T., Noren, A., Olago, D., Owen, R.B., Potts, R., Reed, K., Renaut, R., Schäbitz, F., Tiercelin, J.-J., Trauth, M.H., Wynn, J., Ivory, S., Brady, K., Grady, R., Rodysill, J., Githiri, J., Russell, J., Foerster, V., Dommain, R., Rucina, S., Deocampo, D., Russell, J., Billingsley, A., Beck, C., Dorenbeck, G., Dullo, L., Feary, D., Garello, D., Gromig, R., Johnson, T., Junginger, A., Karanja, M., Kimburi, E., Mbuthia, A., McCartney, T., McNulty, E., Muiruri, V., Nambiro, E., Negash, E.W., Njagi, D., Wilson, J.N., Rabideaux, N., Raub, T., Sier, M.J., Smith, P., Urban, J., Warren, M., Yadeta, M., Yost, C., Zinaye, B., 2016. The Hominin Sites and Paleolakes Drilling Project: inferring the environmental context of human evolution from eastern African rift lake deposits. *Sci. Drill.* 21, 1–16. <https://doi.org/10.5194/sd-21-1-2016>
- Coppens, Y., 1994. East side story: the origin of humankind. *Sci. Am.* 270, 88–95.
- Curry, B.B., 2007. The late-glacial and early Holocene geology, paleoecology and paleohydrology of the Brewster Creek Site, a proposed wetland restoration site, Pratt's Wayne Woods Forest Preserve and James "Pate" Philip State Park, Bartlett, Illinois. Circ. 571 Ill. Dep. Nat. Resour. Ill. State Geol. Surv.
- Cuthbert, M.O., Ashley, G.M., 2014. A Spring Forward for Hominin Evolution in East Africa. *PLoS ONE* 9, e107358. <https://doi.org/10.1371/journal.pone.0107358>
- Cuthbert, M.O., Gleeson, T., Moosdorf, N., Befus, K.M., Schneider, A., Hartmann, J., Lehner, B., 2019. Global patterns and dynamics of climate–groundwater interactions. *Nat. Clim. Change* 9, 137–141. <https://doi.org/10.1038/s41558-018-0386-4>
- Cuthbert, M.O., Gleeson, T., Reynolds, S.C., Bennett, M.R., Newton, A.C., McCormack, C.J., Ashley, G.M., 2017. Modelling the role of groundwater hydro-refugia in East African hominin evolution and dispersal. *Nat. Commun.* 8, 15696.
- Dale, I.R., Greenway, P.J., 1961. *Kenya Trees and Shrubs*. Hatchards, London.
- Daood, A., 2018. *Pollen Grains Recognition Based on Computer Vision Methods*. Florida Institute of Technology, Melbourne, Florida, USA.
- Daood, A., Ribeiro, E., Bush, M., 2018. Sequential Recognition of Pollen Grain Z-Stacks by Combining CNN and RNN, in: The Thirty-First International Florida Artificial

- Intelligence Research Society Conference. Presented at the The Thirty-First International Florida Artificial Intelligence Research Society Conference, Melbourne, Florida, USA, pp. 8–13.
- Dawson, J.B., 2008. The Gregory Rift Valley and Neogene-recent Volcanoes of Northern Tanzania. Geological Society of London.
- de Garidel-Thoron, T., Marchant, R., Tetard, M., Adebayo, M., Gally, Y., 2020. Automated recognition and picking of foraminifera using the MiSo (microfossil sorter) prototype. EGU Gen. Assem. Conf. Abstr. 18067.
- de Heinzelin, J., 1983. The Omo group: archives of the international Omo Research Expedition. Musee royal de l'Afrique centrale.
- deMenocal, P., 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth Planet. Sci. Lett.* 220, 3–24.
- Demissew, S., Cribb, P., Rasmussen, F., 2004. Field Guide to Ethiopian Orchids. Royal Botanic Gardens, Kew.
- Deocampo, D.M., 2005. Evaporative evolution of surface waters and the role of aqueous CO<sub>2</sub> in magnesium silicate precipitation: Lake Eyasi and Ngorongoro Crater, northern Tanzania. *South Afr. J. Geol.* 108, 493–504. <https://doi.org/10.2113/108.4.493>
- Deocampo, D.M., 2002. Sedimentary Processes and Lithofacies in Lake-Margin Groundwater-Fed Wetlands in East Africa, in: *Sedimentation in Continental Rifts*. Society for Sedimentary Geology, Tulsa, Oklahoma, USA, pp. 295–308.
- Diro, G.T., Black, E., Grimes, D.I.F., 2008. Seasonal forecasting of Ethiopian spring rains. *Meteorol. Appl.* 15, 73–83. <https://doi.org/10.1002/met.63>
- Diro, G.T., Grimes, D.I.F., Black, E., O'Neill, A., Pardo-Iguzquiza, E., 2009. Evaluation of reanalysis rainfall estimates over Ethiopia. *Int. J. Climatol. J. R. Meteorol. Soc.* 29, 67–78.
- Diro, G.T., Tonazzzo, T., Shaffrey, L., 2011. Ethiopian Rainfall in Climate Models, in: Williams, C.J.R., Kniveton, D.R. (Eds.), *African Climate and Climate Change: Physical, Social and Political Perspectives, Advances in Global Change Research*. Springer Netherlands, Dordrecht, pp. 51–69. [https://doi.org/10.1007/978-90-481-3842-5\\_3](https://doi.org/10.1007/978-90-481-3842-5_3)
- Dolnicar, S., Grabler, K., Mazanec, J.A., 1999. A tale of three cities: perceptual charting for analyzing destination images, in: Woodside, A., Mazanec, J., Crouch, G.I., Oppermann, M., Sakai, M.Y. (Eds.), *Consumer Psychology of Tourism, Hospitality and Leisure*. CABI Publishing, New York (USA), pp. 39–62.
- Driese, S., Ashley, G., Li, Z., Hover, V., Owen, R., 2004. Possible Late Holocene equatorial palaeoclimate record based upon soils spanning the Medieval Warm Period and Little Ice Age, Loboi Plain, Kenya. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 213, 231–250. <https://doi.org/10.1016/j.palaeo.2004.07.009>
- Dunbar, R.I.M., Dunbar, E.P., 1974. Ecological Relations and Niche Separation between Sympatric Terrestrial Primates in Ethiopia. *Folia Primatol. (Basel)* 21, 36–60. <https://doi.org/10.1159/000155595>
- Ebinger, C., Djomani, Y.P., Mbede, E., Foster, A., Dawson, J.B., 1997. Rifting Archaean lithosphere: the Eyasi-Manyara-Natron rifts, East Africa. *J. Geol. Soc.* 154, 947–960. <https://doi.org/10.1144/gsjgs.154.6.0947>
- Edwards, D., 1983. A broad-scale structural classification of vegetation for practical purposes. *Bothalia* 14, 705–712.
- El-Moslimany, A.P., 1983. History of climate and vegetation in the Eastern Mediterranean and the Middle East from the Pleniglacial to the mid-Holocene. University of Washington, Washington, USA.

- Faith, J.T., Rowan, J., Du, A., 2019. Early hominins evolved within non-analog ecosystems. *Proc. Natl. Acad. Sci.* 116, 21478–21483. <https://doi.org/10.1073/pnas.1909284116>
- Fan, Y., Duffy, C.J., Oliver, D.S., 1997. Density-driven groundwater flow in closed desert basins: field investigations and numerical experiments. *J. Hydrol.* 196, 139–184. [https://doi.org/10.1016/S0022-1694\(96\)03292-1](https://doi.org/10.1016/S0022-1694(96)03292-1)
- Fernandez-Delgado, M., Carrion, P., Cernadas, E., Galvez, J.F., 2003. Improved Classification of Pollen Texture Images Using Svm and Mlp, in: 3rd IASTED International Conference on Visualization, Imaging and Image Processing. Presented at the 3rd IASTED International Conference on Visualization, Imaging and Image Processing, Benalmadena, Spain.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. <https://doi.org/10.1002/joc.5086>
- Flenley, J.R., 1968. The problem of pollen recognition. CSIRO, Canberra, Australia.
- Foster, A., Ebinger, C., Mbede, E., Rex, D., 1997. Tectonic development of the northern Tanzania sector of the East African Rift System. *J. Geol. Soc.* 154, 689–700. <https://doi.org/10.1144/gsjgs.154.4.0689>
- Frey, W., Kürschner, H., 1989. A VI 1 Vorderer Orient. Vegetation.
- Friis, I., Demissew, S., Breugel, P. van, 2010a. Atlas of the potential vegetation of Ethiopia,. The Royal Danish Academy of Sciences and Letters. ed. Det Kongelige Danske Videnskabernes Selskab, Copenhagen.
- Friis, I., Demissew, S., Van Breugel, P., 2010b. Atlas of the potential vegetation of Ethiopia. Det Kongelige Danske Videnskabernes Selskab.
- Frostick, L.E., 1997. Chapter 9 The east african rift basins, in: Selley, R.C. (Ed.), *Sedimentary Basins of the World, African Basins*. Elsevier, pp. 187–209. [https://doi.org/10.1016/S1874-5997\(97\)80012-3](https://doi.org/10.1016/S1874-5997(97)80012-3)
- García, N.M., Chaves, V.A.E., Briceño, J.C., Travieso, C.M., 2012. Pollen Grains Contour Analysis on Verification Approach, in: International Conference on Hybrid Artificial Intelligence Systems. Presented at the International Conference on Hybrid Artificial Intelligence Systems, Salamanca, Spain, pp. 521–532.
- Garello, D.I., 2019. Tephrostratigraphy of Pliocene Drill Cores from Kenya and Ethiopia, and Pleistocene Exposures in the Ledi-Geraru Research Project Area, Ethiopia:Geological Context for the Evolution of Australopithecus and Homo (Ph.D.). Ann Arbor, United States.
- Garnier, A., Neumann, K., Eichhorn, B., Lespez, L., 2013. Phytolith taphonomy in the middle- to late-Holocene fluvial sediments of Ounjougou (Mali, West Africa). *The Holocene* 23, 416–431. <https://doi.org/10.1177/0959683612463102>
- Geheb, K., Abebe, Y.D., 2003. Wetlands of Ethiopia: Proceedings of a Seminar on the Resources and Status of Ethiopia's Wetlands. IUCN.
- Giraud, T., Lambert, N., 2017. Reproducible Cartography, in: Peterson, M.P. (Ed.), *Advances in Cartography and GIScience, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham, pp. 173–183. [https://doi.org/10.1007/978-3-319-57336-6\\_13](https://doi.org/10.1007/978-3-319-57336-6_13)
- Goman, M., Ashley, G.M., Owen, R.B., Hover, V.C., Maharjan, D.K., 2017. Late Holocene Environmental Reconstructions from Lake Solai, Kenya. *Prof. Geogr.* 69, 438–454. <https://doi.org/10.1080/00330124.2016.1266948>
- Gonçalves, A.B., Souza, J.S., Silva, G.G. da, Cereda, M.P., Pott, A., Naka, M.H., Pistori, H., 2016. Feature Extraction and Machine Learning for the Classification of Brazilian

- Savannah Pollen Grains. PLOS ONE 11, e0157044.  
<https://doi.org/10.1371/journal.pone.0157044>
- Greenway, P.J., Vesey-Fitzgerald, D.F., 1969. Vegetation of Lake Manyara National Park. *J. Ecol.* 57, 127–149. <https://doi.org/10.2307/2258212>
- Guiot, J., 1990. Methodology of the last climatic cycle reconstruction in France from pollen data. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 80, 49–69. [https://doi.org/10.1016/0031-0182\(90\)90033-4](https://doi.org/10.1016/0031-0182(90)90033-4)
- Haile-Selassie, Y., Deino, A., Saylor, B., Umer, M., Latimer, B., 2007. Preliminary geology and paleontology of new hominid-bearing Pliocene localities in the central Afar region of Ethiopia. *Anthropol. Sci.* 115, 215–222. <https://doi.org/10.1537/ase.070426>
- Haile-Selassie, Y., Gibert, L., Melillo, S.M., Ryan, T.M., Alene, M., Deino, A., Levin, N.E., Scott, G., Saylor, B.Z., 2015. New species from Ethiopia further expands Middle Pliocene hominin diversity. *Nature* 521, 483.
- Haile-Selassie, Y., Latimer, B.M., Alene, M., Deino, A.L., Gibert, L., Melillo, S.M., Saylor, B.Z., Scott, G.R., Lovejoy, C.O., 2010a. An early *Australopithecus afarensis* postcranium from Woranso-Mille, Ethiopia. *Proc. Natl. Acad. Sci.* 107, 12121–12126. <https://doi.org/10.1073/pnas.1004527107>
- Haile-Selassie, Y., Melillo, S.M., Vazzana, A., Benazzi, S., Ryan, T.M., 2019. A 3.8-million-year-old hominin cranium from Woranso-Mille, Ethiopia. *Nature* 573, 214–219. <https://doi.org/10.1038/s41586-019-1513-8>
- Haile-Selassie, Y., Saylor, B.Z., Deino, A., Alene, M., Latimer, B.M., 2010b. New hominid fossils from Woranso-Mille (Central Afar, Ethiopia) and taxonomy of early *Australopithecus*. *Am. J. Phys. Anthropol.* n/a-n/a. <https://doi.org/10.1002/ajpa.21159>
- Haile-Selassie, Y., Saylor, B.Z., Deino, A., Levin, N.E., Alene, M., Latimer, B.M., 2012. A new hominin foot from Ethiopia shows multiple Pliocene bipedal adaptations. *Nature* 483, 565–569. <https://doi.org/10.1038/nature10922>
- Haile-Selassie, Y., Suwa, G., White, T.D., 2004. Late Miocene Teeth from Middle Awash, Ethiopia, and Early Hominid Dental Evolution. *Science* 303, 1503–1505. <https://doi.org/10.1126/science.1092978>
- Han, L., Xie, Y., 2018. Local Decimal Pattern for Pollen Image Recognition, in: Artificial Neural Networks and Machine Learning – ICANN 2018. Presented at the Artificial Neural Networks and Machine Learning – ICANN 2018, Rhodes, Greece, pp. 47–55.
- Harmand, S., Lewis, J.E., Feibel, C.S., Lepre, C.J., Prat, S., Lenoble, A., Boës, X., Quinn, R.L., Brenet, M., Arroyo, A., Taylor, N., Clément, S., Daver, G., Brugal, J.-P., Leakey, L., Mortlock, R.A., Wright, J.D., Lokorodi, S., Kirwa, C., Kent, D.V., Roche, H., 2015. 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature* 521, 310–315. <https://doi.org/10.1038/nature14464>
- Harrison, S.P., Prentice, I.C., Barboni, D., Kohfeld, K.E., Ni, J., Sutra, J.-P., 2010. Ecophysiological and bioclimatic foundations for a global plant functional classification. *J. Veg. Sci.* 21, 300–317. <https://doi.org/10.1111/j.1654-1103.2009.01144.x>
- Hay, R.L., Kyser, T.K., 2001. Chemical sedimentology and paleoenvironmental history of Lake Olduvai, a Pliocene lake in northern Tanzania. *GSA Bull.* 113, 1505–1521. [https://doi.org/10.1130/0016-7606\(2001\)113<1505:CSAPHO>2.0.CO;2](https://doi.org/10.1130/0016-7606(2001)113<1505:CSAPHO>2.0.CO;2)
- Hély, C., Bremond, L., Alleaume, S., Smith, B., Sykes, M.T., Guiot, J., 2006. Sensitivity of African biomes to changes in the precipitation regime. *Glob. Ecol. Biogeogr.* 15, 258–270.

- Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12, e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Herlocker, D.J., Dirschl, H.J., 1972. Vegetation of the Ngorongoro Conservation Area, Tanzania (No. 19). Canadian Wildlife Service.
- Holdridge, L.R., 1967a. Life zone ecology, Rev. ed. ed. Tropical ScienceCenter, San Jose (Costa Rica).
- Holdridge, L.R., 1967b. Life zone ecology, Tropical Science Center. ed. Tropical Science Center, San José, Costa Rica.
- Holt, K., Allen, G., Hodgson, R., Marsland, S., Flenley, J., 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Rev. Palaeobot. Palynol.* 167, 175–183. <https://doi.org/10.1016/j.revpalbo.2011.08.006>
- Holt, K.A., Bennett, K.D., 2014. Principles and methods for automated palynology. *New Phytol.* 203, 735–742. <https://doi.org/10.1111/nph.12848>
- Hughes, R.H., 1992. A Directory of African Wetlands. IUCN.
- Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Iliffe, J., 1987. The African Poor: A History, African Studies. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511584121>
- Izumi, K., Lézine, A.-M., 2016. Pollen-based biome reconstructions over the past 18,000 years and atmospheric CO<sub>2</sub> impacts on vegetation in equatorial mountains of Africa. *Quat. Sci. Rev.* 152, 93–103.
- Johanson, D., 2017. The paleoanthropology of Hadar, Ethiopia. *Comptes Rendus Palevol* 16, 140–154.
- Johanson, D.C., Taieb, M., 1976. Plio—Pleistocene hominid discoveries in Hadar, Ethiopia. *Nature* 260, 293–297. <https://doi.org/10.1038/260293a0>
- Johanson, D.C., White, T.D., Coppens, Y., 1978. A new species of the genus *Australopithecus* (Primates: Hominidae) from the Pliocene of eastern Africa. *Kirtlandia* 28, 1–14.
- Johnson, T.C., Barry, S.L., Chan, Y., Wilkinson, P., 2001. Decadal record of climate variability spanning the past 700 yr in the Southern Tropics of East Africa. *Geology* 29, 83–86.
- Johnson, T.C., McCave, I.N., 2008. Transport mechanism and paleoclimatic significance of terrigenous silt deposited in varved sediments of an African rift lake. *Limnol. Oceanogr.* 53, 1622–1632.
- Jolly, D., Prentice, I.C., Bonnefille, R., Ballouche, A., Bengo, M., Brenac, P., Buchet, G., Burney, D., Cazet, J.-P., Cheddadi, R., 1998. Biome reconstruction from pollen and plant macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years. *J. Biogeogr.* 25, 1007–1027.
- Kaya, Y., Pinar, S.M., Erez, M.E., Fidan, M., 2013. An expert classification system of pollen of *Onopordum* using a rough set approach. *Rev. Palaeobot. Palynol.* 189, 50–56. <https://doi.org/10.1016/j.revpalbo.2012.11.004>
- Khanzhina, N., Putin, E., Filchenkov, A., Zamyatina, E., 2018. Pollen grain recognition using convolutional neural network, in: Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018). Presented at the Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018), Bruges, Belgium, pp. 409–414.

- Kobayashi, T., Tateishi, R., Alsaideh, B., Sharma, R., Wakaizumi, T., Miyamoto, D., Bai, X., Long, B., Gegentana, G., Maitiniyazi, A., Cahyana, D., Haireti, A., Morifushi, Y., Abake, G., Pratama, R., Zhang, N., Alifu, Z., Shirahata, T., Mi, L., Iizuka, K., Yusupujang, A., Rinawan, F., Bhattacharai, R., Phong, D., 2017. Production of Global Land Cover Data – GLCNMO2013. *J. Geogr. Geol.* 9, p1. <https://doi.org/10.5539/jgg.v9n3p1>
- Kong, S., Punyasena, S., Fowlkes, C., 2016. Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, Nevada, USA, pp. 1–10.
- Lamarck, J.B., 1809. Zoological Philosophy: An Exposition with Regard to the Natural History of Animals, Dentu. ed. Paris, France.
- Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M., Levrard, B., 2004. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* 428, 261–285. <https://doi.org/10.1051/0004-6361:20041335>
- Latif, M., Dommenget, D., Dima, M., Grötzner, A., 1999. The role of Indian Ocean sea surface temperature in forcing East African rainfall anomalies during December–January 1997/98. *J. Clim.* 12, 3497–3504.
- Leakey, M.D., 1971. Olduvai Gorge: Volume 3, Excavations in Beds I and II, 1960–1963. Cambridge University Press.
- Leakey, M.D., 1966. A Review of the Oldowan Culture from Olduvai Gorge, Tanzania. *Nature* 210, 462–466. <https://doi.org/10.1038/210462a0>
- Lebamba, J., Ngomanda, A., Vincens, A., Jolly, D., Favier, C., Elenga, H., Bentaleb, I., 2009. Central African biomes and forest succession stages derived from modern pollen data and plant functional types. *Clim. Past* 5.
- Lebart, L., 2000. Contiguity Analysis and Classification, in: Gaul, W., Opitz, O., Schader, M. (Eds.), *Data Analysis: Scientific Modeling and Practical Application, Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg, pp. 233–243. [https://doi.org/10.1007/978-3-642-58250-9\\_19](https://doi.org/10.1007/978-3-642-58250-9_19)
- Lebart, L., Morineau, A., Piron, M., 1995. *Statistique exploratoire multidimensionnelle*. Dunod, Paris.
- Levin, N.E., 2015. Environment and climate of early human evolution. *Annu. Rev. Earth Planet. Sci.* 43, 405–429.
- Lézine, A.-M., 1981. Le Lac Abiyata (Ethiopie) - Palynologie et paléoclimatologie du Quaternaire récent. Université de Bordeaux, Bordeaux, France.
- Lézine, A.-M., Saliège, J.-F., Robert, C., Wertz, F., Inizan, M.-L., 1998. Holocene Lakes from Ramlat as-Sab'atayn (Yemen) Illustrate the Impact of Monsoon Activity in Southern Arabia. *Quat. Res.* 50, 290–299. <https://doi.org/10.1006/qres.1998.1996>
- Li, P., Treloar, W.J., Flenley, J.R., Empson, L., 2004. Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. *J. Quat. Sci.* 19, 755–762. <https://doi.org/10.1002/jqs.874>
- Li, Y., Wang, N., Li, Z., Zhang, C., Zhou, X., 2012. Reworking effects in the Holocene Zhuye Lake sediments: A case study by pollen concentrates AMS 14C dating. *Sci. China Earth Sci.* 55, 1669–1678. <https://doi.org/10.1007/s11430-012-4482-4>
- Lisicki, L.E., Raymo, M.E., 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* 20. <https://doi.org/10.1029/2004PA001071>

- López-Sáez, J.-A., Domínguez-Rodrigo, M., 2009. Palynology of OGS-6a and OGS-7, two new 2.6 Ma archaeological sites from Gona, Afar, Ethiopia: Insights on aspects of Late Pliocene habitats and the beginnings of stone-tool use. *Geobios* 42, 503–511. <https://doi.org/10.1016/j.geobios.2008.12.002>
- Loth, P., Prins, H., 1986. Spatial patterns of landscape and vegetation of Lake Manyara National Park.
- Loth, P.E., Prins, H.H.T., 1986. Spatial patterns of the landscape and vegetation of Lake Manyara National Park. *ITC J.* 2, 115–130.
- Lugo, A.E., Brown, S.L., Dodson, R., Smith, T.S., Shugart, H.H., 1999. The Holdridge life zones of the conterminous United States in relation to ecosystem mapping. *J. Biogeogr.* 26, 1025–1038. <https://doi.org/10.1046/j.1365-2699.1999.00329.x>
- Lüning, S., Galka, M., Danladi, I.B., Adagunodo, T.A., Vahrenholt, F., 2018. Hydroclimate in Africa during the medieval climate anomaly. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 495, 309–322.
- Lyon, B., DeWitt, D.G., 2012. A recent and abrupt decline in the East African long rains. *Geophys. Res. Lett.* 39. <https://doi.org/10.1029/2011GL050337>
- Magill, C.R., Ashley, G.M., Domínguez-Rodrigo, M., Freeman, K.H., 2016a. Dietary options and behavior suggested by plant biomarker evidence in an early human habitat. *Proc. Natl. Acad. Sci.* 113, 2874–2879. <https://doi.org/10.1073/pnas.1507055113>
- Magill, C.R., Ashley, G.M., Domínguez-Rodrigo, M., Freeman, K.H., 2016b. Dietary options and behavior suggested by plant biomarker evidence in an early human habitat. *Proc. Natl. Acad. Sci.* 113, 2874–2879.
- Magill, C.R., Ashley, G.M., Freeman, K.H., 2013a. Water, plants, and early human habitats in eastern Africa. *Proc. Natl. Acad. Sci.* 110, 1175–1180. <https://doi.org/10.1073/pnas.1209405109>
- Magill, C.R., Ashley, G.M., Freeman, K.H., 2013b. Ecosystem variability and early human habitats in eastern Africa. *Proc. Natl. Acad. Sci.* 110, 1167–1174. <https://doi.org/10.1073/pnas.1206276110>
- Mander, L., Baker, S.J., Belcher, C.M., Haselhorst, D.S., Rodriguez, J., Thorn, J.L., Tiwari, S., Urrego, D.H., Wesseln, C.J., Punyasena, S.W., 2014. Accuracy and consistency of grass pollen identification by human analysts using electron micrographs of surface ornamentation. *Appl. Plant Sci.* 2, 1400031. <https://doi.org/10.3732/apps.1400031>
- Marchant, R., Tetard, M., de Garidel-Thoron, T., In Preparation. A System for Classification of Foraminifera Images Using Deep Convolutional Neural Networks.
- Marcos, J.V., Nava, R., Cristóbal, G., Redondo, R., Escalante-Ramírez, B., Bueno, G., Déniz, Ó., González-Porto, A., Pardo, C., Chung, F., Rodríguez, T., 2015. Automated pollen identification using microscopic imaging and texture analysis. *Micron* 68, 36–46. <https://doi.org/10.1016/j.micron.2014.09.002>
- Maslin, M.A., Shultz, S., Trauth, M.H., 2015. A synthesis of the theories and concepts of early human evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20140064. <https://doi.org/10.1098/rstb.2014.0064>
- Maslin, M.A., Trauth, M.H., 2009. Plio-Pleistocene East African pulsed climate variability and its influence on early human evolution, in: The First Humans—Origin and Early Evolution of the Genus Homo. Springer, pp. 151–158.
- McPherron, S.P., Alemseged, Z., Marean, C.W., Wynn, J.G., Reed, D., Geraads, D., Bobe, R., Béarat, H.A., 2010. Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature* 466, 857–860. <https://doi.org/10.1038/nature09248>

- Miehe, S., Miehe, G., 1994. Ericaceous forests and heathlands in the Bale Mountains of South Ethiopia: ecology and man's impact, Stiftung Walderhaltung in Afrika and Bundesforschungsanstalt für Forst und Holzwirtschaft. ed. T. Warnke Verlag, Hamburg.
- Mohammed, M.U., 1992. Paléoenvironnement et paléoclimatologie des derniers millénaires en Ethiopie. Contribution palynologique. University of Aix-Marseille III, Marseille, France.
- Mohammed, M.U., Bonnefille, R., 2002. Late Holocene Climatic Fluctuations and Historical Records of Famine in Ethiopia, in: Droughts, Food and Culture. Springer, pp. 83–94.
- Mohammed, M.U., Lamb, H.F., Bonnefille, R., Lézine, A.-M., Tiercelin, J.-J., Gibert, E., Cazet, J.-P., Watrin, J., 2007. Late pleistocene and holocene vegetation history of the bale mountains, Ethiopia. *Quat. Sci. Rev.* 26, 2229–2246.
- Mohan, J., Stone, J.R., Campisano, C.J., 2016. Three novel species of Bacillariophyta (Diatoms) belonging to Aulacoseira and Lindavia from the Pliocene Hadar Formation, Afar Depression of Ethiopia. *Phytotaxa* 272, 235–247. <https://doi.org/10.11646/phytotaxa.272.4.1>
- Nash, D.J., De Cort, G., Chase, B.M., Verschuren, D., Nicholson, S.E., Shanahan, T.M., Asrat, A., Lézine, A.-M., Grab, S.W., 2016. African hydroclimatic variability during the last 2000 years. *Quat. Sci. Rev.* 154, 1–22.
- Nicholson, S.E., Klotter, D., Dezfuli, A.K., 2012. Spatial reconstruction of semi-quantitative precipitation fields over Africa during the nineteenth century from documentary evidence and gauge data. *Quat. Res.* 78, 13–23.
- Norton, M.K., 2019. A hydrogeologic framework for spring formation in the Ngorongoro Volcanic Highland region (Thesis for the degree of Master). Rutgers University, New Brunswick, New Jersey, USA. <https://doi.org/10.7282/T3ZS2V4M>
- Öberg, H., Norström, E., Malmström Ryner, M., Holmgren, K., Westerberg, L.-O., Risberg, J., Eddudóttir, S.D., Andersen, T.J., Muzuka, A., 2013. Environmental variability in northern Tanzania from AD 1000 to 1800, as inferred from diatoms and pollen in Lake Duluti. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 374, 230–241. <https://doi.org/10.1016/j.palaeo.2013.01.021>
- Oliphant, T.E., 2015. A guide to NumPy, CreateSpace Independent Publishing Platform. ed. North Charleston, USA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pérez, F., Granger, B.E., Hunter, J.D., 2011. Python: An Ecosystem for Scientific Computing. *Comput. Sci. Eng.* 13, 13–21. <https://doi.org/10.1109/MCSE.2010.119>
- Peyron, O., Jolly, D., Bonnefille, R., Vincens, A., Guiot, J., 2000. Climate of East Africa 6000  $^{14}\text{C}$  yr BP as inferred from pollen data. *Quat. Res.* 54, 90–101.
- Pickering, R., 1964. Endulen, Quarter degree sheet 52. Records of the Geological Survey of Tanganyika.
- Prentice, C., Guiot, J., Huntley, B., Jolly, D., Cheddadi, R., 1996. Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Clim. Dyn.* 12, 185–194.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 117–134.

- Prentice, I.C., Webb III, T., 1998. BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records. *J. Biogeogr.* 25, 997–1005.
- Punyasena, S.W., Tcheng, D.K., Wesseln, C., Mueller, P.G., 2012. Classifying black and white spruce pollen using layered machine learning. *New Phytol.* 196, 937–944. <https://doi.org/10.1111/j.1469-8137.2012.04291.x>
- Quade, J., Levin, N., Semaw, S., Stout, D., Renne, P., Rogers, M., Simpson, S., 2004. Paleoenvironments of the earliest stone toolmakers, Gona, Ethiopia. *GSA Bull.* 116, 1529–1544. <https://doi.org/10.1130/B25358.1>
- Rea, R.J.A., 1935. The forest types of vegetation in tanganyika territory. *Emp. For. J.* 14, 202–208.
- Reed, K.E., 2008. Paleoecological patterns at the Hadar hominin site, Afar Regional State, Ethiopia. *J. Hum. Evol.* 54, 743–768. <https://doi.org/10.1016/j.jhevol.2007.08.013>
- Reed, K.E., Bidner, L.R., 2004. Primate communities: Past, present, and possible future. *Am. J. Phys. Anthropol.* 125, 2–39. <https://doi.org/10.1002/ajpa.20153>
- Reed, K.E., Spencer, L.M., Rector, A.L., 2013. Faunal approaches in early hominin paleo-ecology, in: Early Hominin Paleoecology. University Press of Colorado.
- Reynolds, S.C., Marston, C.G., Hassani, H., King, G.C.P., Bennett, M.R., 2016. Environmental hydro-refugia demonstrated by vegetation vigour in the Okavango Delta, Botswana. *Sci. Rep.* 6, 35951. <https://doi.org/10.1038/srep35951>
- Reynolds, S.C., Wilkinson, D.M., Marston, C.G., O'Regan, H.J., 2015. The 'mosaic habitat'concept in human evolution: past and present. *Trans. R. Soc. South Afr.* 70, 57–69.
- Richmond, B.G., Jungers, W.L., 2008. Orrorin tugenensis Femoral Morphology and the Evolution of Hominin Bipedalism. *Science* 319, 1662–1665. <https://doi.org/10.1126/science.1154197>
- Rodriguez-Damian, M., Cernadas, E., Formella, A., Fernandez-Delgado, M., Sa-Otero, P.D., 2006. Automatic detection and classification of grains of pollen based on shape and texture. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 36, 531–542. <https://doi.org/10.1109/TSMCC.2005.855426>
- Roman, D.C., Campisano, C.J., Quade, J., DiMaggio, E.N., Arrowsmith, J.R., Feibel, C., 2008. Composite tephrostratigraphy of the Dikika, Gona, Hadar, and Ledi-Geraru project areas, northern Awash, Ethiopia, in: The Geology of Early Humans in the Horn of Africa. Geological Society of America, Boulder, Colorado, pp. 119–134.
- Ruff, C.B., Burgess, M.L., Ketcham, R.A., Kappelman, J., 2016. Limb Bone Structural Proportions and Locomotor Behavior in A.L. 288-1 ("Lucy"). *PLOS ONE* 11, e0166095. <https://doi.org/10.1371/journal.pone.0166095>
- Russell, J.M., Johnson, T.C., 2007. Little Ice Age drought in equatorial Africa: Intertropical Convergence Zone migrations and El Niño–Southern Oscillation variability. *Geology* 35, 21–24. <https://doi.org/10.1130/G23125A.1>
- Russell, J.M., Johnson, T.C., 2005. A high-resolution geochemical record from Lake Edward, Uganda Congo and the timing and causes of tropical African drought during the late Holocene. *Quat. Sci. Rev.* 24, 1375–1389.
- Russell, J.M., Verschuren, D., Eggermont, H., 2007. Spatial complexity of 'Little Ice Age' climate in East Africa: sedimentary records from two crater lake basins in western Uganda. *The Holocene* 17, 183–193. <https://doi.org/10.1177/0959683607075832>
- Ryner, M., Holmgren, K., Taylor, D., 2008. A record of vegetation dynamics and lake level changes from Lake Emakat, northern Tanzania, during the last c. 1200 years. *J. Paleolimnol.* 40, 583–601. <https://doi.org/10.1007/s10933-007-9184-0>

- Saji, N.H., Goswami, B.N., Vinayachandran, P.N., Yamagata, T., 1999. A dipole mode in the tropical Indian Ocean. *Nature* 401, 360–363.
- Saylor, B.Z., Gibert, L., Deino, A., Alene, M., Levin, N.E., Melillo, S.M., Peaple, M.D., Feakins, S.J., Bourel, B., Barboni, D., 2019. Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia. *Nature* 573, 220–224.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>
- Semaw, S., Rogers, M.J., Quade, J., Renne, P.R., Butler, R.F., Dominguez-Rodrigo, M., Stout, D., Hart, W.S., Pickering, T., Simpson, S.W., 2003. 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *J. Hum. Evol.* 45, 169–177. [https://doi.org/10.1016/S0047-2484\(03\)00093-9](https://doi.org/10.1016/S0047-2484(03)00093-9)
- Semaw, S., Simpson, S.W., Quade, J., Renne, P.R., Butler, R.F., McIntosh, W.C., Levin, N., Dominguez-Rodrigo, M., Rogers, M.J., 2005. Early Pliocene hominids from Gona, Ethiopia. *Nature* 433, 301–305. <https://doi.org/10.1038/nature03177>
- Sevillano, V., Aznarte, J.L., 2018. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. *PloS One* 13, e0201807. <https://doi.org/10.1371/journal.pone.0201807>
- Shilling, A.M., 2013. Characterization of groundwater discharge sites using remote sensing and wetland cores, Lake Eyasi basin, Tanzania (Thesis for the degree of Master). Rutgers University, New Brunswick, New Jersey, USA. <https://doi.org/10.7282/T3ZS2V4M>
- Sikes, N.E., Ashley, G.M., 2007. Stable isotopes of pedogenic carbonates as indicators of paleoecology in the Plio-Pleistocene (upper Bed I), western margin of the Olduvai Basin, Tanzania. *J. Hum. Evol., African Paleoclimate and Human Evolution* 53, 574–594. <https://doi.org/10.1016/j.jhevol.2006.12.008>
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* 1409–1556.
- Sobol, M.K., Finkelstein, S.A., 2018. Predictive pollen-based biome modeling using machine learning. *PLOS ONE* 13, e0202214. <https://doi.org/10.1371/journal.pone.0202214>
- Spoor, F., Leakey, M.G., O'Higgins, P., 2016. Middle Pliocene hominin diversity: *Australopithecus deyiremeda* and *Kenyanthropus platyops*. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150231. <https://doi.org/10.1098/rstb.2015.0231>
- Stager, J.C., Ryves, D., Cumming, B.F., Meeker, L.D., Beer, J., 2005. Solar variability and the levels of Lake Victoria, East Africa, during the last millennium. *J. Paleolimnol.* 33, 243–251.
- Stewart, K.M., 2014. Environmental change and hominin exploitation of C4-based resources in wetland/savanna mosaics. *J. Hum. Evol.* 77, 1–16.
- Suchéras-Marx, B., Escarguel, G., Ferreira, J., Hammer, Ø., 2019. Statistical confidence intervals for relative abundances and abundance-based ratios: Simple practical solutions for an old overlooked question. *Mar. Micropaleontol.* 151, 101751. <https://doi.org/10.1016/j.marmicro.2019.101751>
- Taieb, M., Tiercelin, J.J., 1979. Sedimentation pliocene et paleoenvironnements de rift; exemple de la formation a Hominides d'Hadar (Afar, Ethiopie). *Bull. Société Géologique Fr. S7-XXI*, 243–253. <https://doi.org/10.2113/gssgbull.S7-XXI.3.243>

- Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., Beaufort, L., 2020. A new automated radiolarian image acquisition, stacking, processing, segmentation, and identification workflow. *Clim. Past Discuss.* 1–23. <https://doi.org/10.5194/cp-2020-76>
- Thompson, L.G., Mosley-Thompson, E., Davis, M.E., Henderson, K.A., Brecher, H.H., Zagorodnov, V.S., Mashiotta, T.A., Lin, P.-N., Mikhaleko, V.N., Hardy, D.R., Beer, J., 2002. Kilimanjaro Ice Core Records: Evidence of Holocene Climate Change in Tropical Africa. *Science* 298, 589–593. <https://doi.org/10.1126/science.1073198>
- Tierney, J.E., Smerdon, J.E., Anchukaitis, K.J., Seager, R., 2013. Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. *Nature* 493, 389–392.
- Trabucco, A., Zomer, R.J., 2019. Global High-Resolution Soil-Water Balance. [WWW Document]. Glob. High-Resolut. Soil-Water Balance. URL <https://doi.org/10.6084/m9.figshare.7707605.v3> (accessed 12.6.19).
- Trauth, M.H., Maslin, M.A., Deino, A.L., Strecker, M.R., Bergner, A.G.N., Dühnforth, M., 2007. High- and low-latitude forcing of Plio-Pleistocene East African climate and human evolution. *J. Hum. Evol., African Paleoclimate and Human Evolution* 53, 475–486. <https://doi.org/10.1016/j.jhevol.2006.12.009>
- Treloar, W.J., Taylor, G.E., Flenley, J.R., 2004. Towards automation of palynology 1: analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images. *J. Quat. Sci.* 19, 745–754. <https://doi.org/10.1002/jqs.871>
- van Breugel, P., Breugel, P. van, Kindt, R., Lillesø, J.-P.B., Bingham, M., Demissew, S., Dudley, C., Friis, I., Gachathi, F., Kalema, J., Mbago, F.M., Moshi, H.N., Mulumba, J., Namaganda, M., Ndangalasi, H.J., Ruffo, C.K., Védaste, M., Jamnadass, R., Graudal, L., 2015. Potential Natural Vegetation Map of Eastern Africa (Burundi, Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda and Zambia). Version 2.0.
- van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. Scikit-image: image processing in Python. *PeerJ* 2, e453. <https://doi.org/10.7717/peerj.453>
- van Holstein, L., Foley, R.A., 2017. Hominin Evolution, in: Shackelford, T.K., Weekes-Shackelford, V.A. (Eds.), *Encyclopedia of Evolutionary Psychological Science*. Springer International Publishing, Cham, Switzerland, pp. 1–22. [https://doi.org/10.1007/978-3-319-16999-6\\_3416-1](https://doi.org/10.1007/978-3-319-16999-6_3416-1)
- Verschuren, D., 2004. Decadal and century-scale climate variability in tropical Africa during the past 2000 years, in: Battarbee, R.W., Gasse, F., Stickley, C.E. (Eds.), *Past Climate Variability through Europe and Africa, Developments in Paleoenvironmental Research*. Springer Netherlands, Dordrecht, pp. 139–158. [https://doi.org/10.1007/978-1-4020-2121-3\\_8](https://doi.org/10.1007/978-1-4020-2121-3_8)
- Verschuren, D., 2001. Reconstructing fluctuations of a shallow East African lake during the past 1800 yrs from sediment stratigraphy in a submerged crater basin. *J. Paleolimnol.* 25, 297–311. <https://doi.org/10.1023/A:1011150300252>
- Verschuren, D., Damste, J.S.S., Moernaut, J., Kristen, I., Blaauw, M., Fagot, M., Haug, G.H., 2009. Half-precessional dynamics of monsoon rainfall near the East African Equator. *Nature* 462, 637–641. <https://doi.org/10.1038/nature08520>
- Verschuren, D., Laird, K.R., Cumming, B.F., 2000. Rainfall and drought in equatorial east Africa during the past 1,100 years. *Nature* 403, 410–414.
- Villmoare, B., Kimbel, W.H., Seyoum, C., Campisano, C.J., DiMaggio, E.N., Rowan, J., Braun, D.R., Arrowsmith, J.R., Reed, K.E., 2015. Early Homo at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. *Science* 347, 1352–1355. <https://doi.org/10.1126/science.aaa1343>

- Vincens, A., Bremond, L., Brewer, S., Buchet, G., Dussouillez, P., 2006a. Modern pollen-based biome reconstructions in East Africa expanded to southern Tanzania. *Rev. Palaeobot. Palynol.* 140, 187–212.
- Vincens, A., Bremond, L., Brewer, S., Buchet, G., Dussouillez, P., 2006b. Modern pollen-based biome reconstructions in East Africa expanded to southern Tanzania. *Rev. Palaeobot. Palynol.* 140, 187–212.
- Vincens, A., Lézine, A.-M., Buchet, G., Lewden, D., Le Thomas, A., 2007. African pollen database inventory of tree and shrub pollen types. *Rev. Palaeobot. Palynol.* 145, 135–141. <https://doi.org/10.1016/j.revpalbo.2006.09.004>
- Ward, C.V., 2002. Interpreting the posture and locomotion of *Australopithecus afarensis*: Where do we stand? *Am. J. Phys. Anthropol.* 119, 185–215. <https://doi.org/10.1002/ajpa.10185>
- Webster, P.J., Moore, A.M., Loschnigg, J.P., Leben, R.R., 1999. Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature* 401, 356–360.
- White, T.D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C.O., Suwa, G., WoldeGabriel, G., 2009. *Ardipithecus ramidus* and the Paleobiology of Early Hominids. *Science* 326, 64–64, 75–86. <https://doi.org/10.1126/science.1175802>
- White, T.D., WoldeGabriel, G., Asfaw, B., Ambrose, S., Beyene, Y., Bernor, R.L., Boisserie, J.-R., Currie, B., Gilbert, H., Haile-Selassie, Y., Hart, W.K., Hlusko, L.J., Howell, F.C., Kono, R.T., Lehmann, T., Louchart, A., Lovejoy, C.O., Renne, P.R., Saegusa, H., Vrba, E.S., Wesselman, H., Suwa, G., 2006. Asa Issie, Aramis and the origin of *Australopithecus*. *Nature* 440, 883–889. <https://doi.org/10.1038/nature04629>
- WoldeGabriel, G., Ambrose, S.H., Barboni, D., Bonnefille, R., Bremond, L., Currie, B., DeGusta, D., Hart, W.K., Murray, A.M., Renne, P.R., 2009. The geological, isotopic, botanical, invertebrate, and lower vertebrate surroundings of *Ardipithecus ramidus*. *Science* 326, 65–65e5.
- Wood, B., 2011. Ledi Geraru study area, in: Wiley-Blackwell Encyclopedia of Human Evolution. Wiley-Blackwell, New York, USA, p. 878.
- Wynn, J.G., Alemseged, Z., Bobe, R., Geraads, D., Reed, D., Roman, D.C., 2006. Geological and palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia. *Nature* 443, 332–336. <https://doi.org/10.1038/nature05048>
- Wynn, J.G., Roman, D.C., Alemseged, Z., Reed, D., Geraads, D., Munro, S., 2008. Stratigraphy, depositional environments, and basin structure of the Hadar and Busidima Formations at Dikika, Ethiopia, in: The Geology of Early Humans in the Horn of Africa. Geological Society of America. [https://doi.org/10.1130/2008.2446\(04\)](https://doi.org/10.1130/2008.2446(04))
- Zhao, M., Heinsch, F.A., Nemani, R.R., Running, S.W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* 95, 164–176. <https://doi.org/10.1016/j.rse.2004.12.011>
- Zheng, A., 2015. Evaluating Machine Learning Models. O'Reilly Media, Sebastopol, Ukrainian/Russia.

# Table des figures

- Figure 1.1** - Répartition chronologique et spatiale des espèces d'Hominini adaptée d'après Maslin et al. (2015) et complétée avec Bobe et al. (2020), Behrensmeyer et Reed (2013), Campisano et al. (2017) et Levin (2015). Elle se base sur les dates de premières et dernières occurrences..... 12
- Figure 1.2** - Développement local de forêts azonales le long de la rivière dans une végétation régionale et zonale de formations arbustives à *Acacia-Commiphora* (moyenne vallée de l'Awash, Ethiopie). Pourcentage de couvert arboré issu de Global Maps (année 2008,  $0.46 \times 0.46$  km, Kobayashi et al., 2017)..... 16
- Figure 2.1** – Distribution of pollen samples from this study. The red circles are the old samples and the white squares are the new samples..... 28
- Figure 2.2** - Rainfall patterns for Ethiopia based on figures from Diro et al. (2008). Annual cycle for the homogenous rainfall zones of Ethiopia. The annual cycle was averaged over the 35 year period (1969–2003). The annual cycle in broken lines represent stations added to the zone after considering the cross correlation criteria. The solid lines represent stations which are retained in the zones after considering the cross correlation criteria. Different symbols indicate different stations in the zone. The pollen samples used in this study are positioned on the map with their post-hoc inference of regional vegetation and information on their edaphic context (see Table 2.1 for abbreviations)..... 29
- Figure 2.3** - Present-day distribution of the vegetation types in the climate domain of Ethiopia. A-H, Climate domain of Ethiopia (grey shading) with each of the 53381 dots representing  $4.63 \times 4.63$  km) defined by mean annual precipitation (mm/yr) (WordClim2, 1970-2000, Fick and Hijmans, 2017). A-G, Climatic domain occupied by the main vegetation types of Ethiopia obtained using the georeferenced map of vegetation types and sub-types over Eastern Africa produced under the Vegetation and Climate change in Eastern Africa (VECEA) project (van Breugel et al., 2015). A: Afroalpine vegetation ; Bd: *Acacia-Commiphora* deciduous bushland and thicket ; D: Desert ; E: Ericaceous belt ; Fa: Afromontane rain forest ; Fb/Wd: Complex of afromontane undifferentiated forest with wooded grasslands and evergreen or semi-evergreen bushland and thicket ; S: Semi-desert grassland and shrubland ; Wcd: Dry *Combretum* wooded grassland (Wcd) ; We: Upland *Acacia* wooded grassland. H, Position of our 249 surface samples within the present-day climatic domain of Ethiopia and according to the post-hoc inference of regional vegetation (Table 2.1)..... 32
- Figure 2.4** - Comparison of the corrected geographical coordinates for the pollen samples E010 to E026 (orange square) with the original geographical coordinates (white square)..... 34
- Figure 2.5** - Range of score values significant (cyan) and non-significant (grey) for each vegetation type and sub-type with a confidence interval of 95%. In brackets,

vegetation types or sub-types with overlapping significant scores. For abbreviations, see Table 2.1.....	42
<b>Figure 2.6</b> - Neighborly relationships between the <i>k</i> -mean groups. Arrows and red numbers indicate the percentage of neighboring samples (eg., 27% of the Bd-spring 2 samples are closest neighbors to the Bd-spring 1 samples and 29% of the Bd-spring 1 samples are closest neighbors to the Bd-spring 2 samples). The length of the arrows is inversely proportional to the proximity of groups. Values below 25% have not been indicated. For abbreviations, see Table 2.1.....	45
<b>Figure 2.7</b> - Ranges of mean annual temperature by mean annual precipitation and mean temperature of warmest quarter by mean precipitation of driest quarter for the post-hoc inferred vegetation represented by the samples (WordClim2, 1970-2000, 0.93x0.93 km, Fick and Hijmans, 2017) with a confidence interval of 95%. For abbreviations, see Table 1.....	48
<b>Figure 2.8</b> - Range of Percent tree cover for the post-hoc inferred vegetation represented by the samples (Global Maps, year 2008, 0.46x0.46 km, Kobayashi et al., 2017) with a confidence interval of 95%. For abbreviations, see Table 1.....	49
<b>Figure 2.9</b> - Map of the modern pollen samples used in this study showing the vegetation types and sub-types inferred from our new approach. Cross, diamond and star symbols distinguish soil, river and spring samples (see Table 2.1 for abbreviations). (see Table 2.1 for abbreviations). .....	50
<b>Figure 2.10</b> - Rainfall patterns with mean monthly precipitation (mm/month) based on our 249 samples for each identified vegetation type. The red area indicates the <30 mm/month precipitation threshold that is used to consider a month dry. For abbreviations, see Table 2.1.....	52
<b>Figure 3.1</b> - Pollen sampling areas by locality in the Northern Awash Basin (Ethiopia).	66
<b>Figure 3.2</b> - Chronological and spatial distribution of Hominini species in the northern part of the Eastern African Rift (Ethiopia and Northern Kenya) between 4.3 Ma and 2.9 based on Bobe et al. (2020), Behrensmeyer et Reed (2013), Campisano et al. (2017) and , Haile-Selassie et al. (2012). .....	67
<b>Figure 3.3</b> - Position of the analyzed levels in the NAO-1B and NAW-1A cores with the degree of pollen preservation. ....	70
<b>Figure 3.4</b> - Position of the pollen samples in the composite stratigraphy of the Hadar and Busidima Formations with respect to time scales. MMB assemblage: Maka'amitalu Basin assemblage ; KH-7: KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-3: Denen Dora Member sub-member 3 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; SHT: Sidi Hakoma Tuff ; UBM: Upper Basal Member ; BBM: Base of Basal Member. ....	78
<b>Figure 3.5</b> – Pollen-inferred climatic parameters and percent of tree cover (confidence interval of 95% ) based on supplementary data of the Part 1 of this thesis (Annexe table 2.3). MMB assemblage: Maka'amitalu Basin assemblage ; KH-7:KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1:	

KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member. For abbreviations of inferred paleovegetation, see Table 3.2.....80

**Figure 3.6** - Bilan of the evolution of climatic and environmental conditions. KH-7:KadaHadar Member sub-member 7 ;KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member.....81

**Figure 3.7** - Comparison of pollen-inferred climatic parameters for Plant Functional Types (PFT)(Bonnefille et al., 2004), la Best Analog (BA) (Bonnefille et al., 2004), and V and V-river scores methods (Part 1 of this thesis) with LR04 benthic  $\delta^{18}\text{O}$  stack (Lisiecki and Raymo, 2005). The red band represents the temperatures and precipitation measured at the Hadar site based on WordClim 2 (Fick and Hijmans, 2017).....87

**Figure 4.1** - Satellite image of sector around the Ngorongoro Volcanic Highlands. Grey triangles are volcanoes, red circles hominin paleontological sites, and red star is our study site. The approximate contour of paleolake Olduvai is for 1.75 Ma age (based on Barboni, 2014). Lake Emakat is the crater lake at Embagai.....97

**Figure 4.2** - Satellite images of the localized vegetation belt on the northeastern margin of Lake Eyasi with a zoom on the Kisima Nggeda area. Sw: swamp, Wg: wet grassland. The letters A and B are respectively the positions of the photo in Figure 3.A and the photo in Figure 3.B.....98

**Figure 4.3** - Theoretical drawing showing vegetation according to the groundwater level which controls the spatial extent of the hydrophytic (A) and mesophytic (B) zones. Photographs show how the two zones may look like in the area of Kisima Nggeda (photographs taken by D. Barboni and G.M. Ashley in 2011).....101

**Figure 4.4** - Photograph of core KN1A cut in half with age models based on calibrated radiocarbon ages of 14C dating of concentrated pollen grains. Red crosses indicate position of samples for 14C dating.....104

**Figure 4.5** - Pollen diagram with major pollen types and interpretation of the pollen signal. Mimosaceae include *Acacia* in black, *Albizia*-type in grey and *Calpocalyx*-type *letestui* in light grey. Black dots indicate pollen percentages <0.5%. Groundwater discharge estimation is based on percentage of *Typha*, *Hyphaene*-type, and Mimosaceae. The leaf symbol represents peaks of Mimosaceae (*Acacia* or *Albizia*-type).....106

**Figure 4.6** - Synthetic pollen diagrams by ecological zones (undifferentiated excluded, cf, Table S1) and by life cycle/leaf phenology (undifferentiated excluded, cf, Table S1). AP: arboreal plants; Halo: halophyte; Hydro: hydrophyte; Meso: mesophyte; NAP: non-arboreal plants; Xero: xerophyte.....107

<b>Figure 4.7</b> - Spring activity of Kisima Ngeda based on pollen percentages of <i>Typha</i> over the past 1200 years with 95% confidence interval (dark gray envelope) compared with other regional proxy records reflecting moisture variations. Light gray stripes indicate the "wet" phases.....	112
<b>Figure 4.8</b> - Kisima Ngeda spring response (wetland expansions) to mega-regional and regional events. <b>(a)</b> Relative abundance of <i>Typha</i> pollen in KN1A core at Kisima Ngeda, Lake Eyasi, N Tanzania. Brown arrow indicates peat inception. <b>(b)</b> Positive (light blue) and extreme positive Indian Ocean Dipole (IOD) events (dark blue) according to the coral $\delta^{18}\text{O}$ record of the southern Mentawai Islands, offshore of Sumatra (Abram et al., 2020). Arrows represent 75 years after positive IOD events. Note that, to date, there is no isotopic record of IOD events older than AD 1240. <b>(c)</b> Amplitude of the first Monte Carlo empirical orthogonal factor (MCEOF) showing a coherent statistical signal of rainfall (as inferred by various proxies) in Eastern Africa over the AD 1300-2000 time period. Kisima Ngeda spring activity exhibits a different signal.....	115
<b>Figure 5.1</b> - Pollen images with automatic z-stack of 128 x 128 pixels for all pollen types of the intact, damaged, and fossil pollen datasets. The name under each image is the final label assigned by the multiple convolutional neural networks (multi-CNNs) with augmentation.....	127
<b>Figure 5.2</b> - Schematic of the classification system based on multiple convolutional neural networks (multi-CNNs) with augmentation. aa: <i>Achyranthes</i> -type <i>aspera</i> ; aj: <i>Aerva</i> -type <i>javanica</i> ; al: <i>Aerva</i> -type <i>lanata</i> ; amar: Amaranthaceae; c: Cyperaceae; ca: <i>Celosia</i> -type <i>argentea</i> ; co: <i>Cyathula</i> -type <i>orthacantha</i> ; dm: <i>Digera</i> -type <i>muricata</i> ; mono: Monocotyledonae; p: Poaceae; uncert: uncertain; indet: indeterminable.	130
<b>Figure 5.3</b> - Results of our classification system (multi-CNNs with augmentation) as a function of pollen dataset and in comparison with other setups. Pollen types of the lowest taxonomic levels: <i>Achyranthes</i> -type <i>aspera</i> , <i>Aerva</i> -type <i>javanica</i> , <i>Aerva</i> -type <i>lanata</i> , <i>Celosia</i> -type <i>argentea</i> , <i>Cyathula</i> -type <i>orthacantha</i> and <i>Digera</i> -type <i>muricata</i> , and Poaceae and Cyperaceae. indet.: indeterminable. The APC% is the percentages based on average per-class.....	132
<b>Figure 5.4</b> - Confusion matrix for the intact (a), damaged (b) and fossil (c) pollen dataset for final classes. aa: <i>Achyranthes</i> -type <i>aspera</i> ; aj: <i>Aerva</i> -type <i>javanica</i> ; al: <i>Aerva</i> -type <i>lanata</i> ; amar: Amaranthaceae; c: Cyperaceae; ca: <i>Celosia</i> <i>argentea</i> ; co: <i>Cyathula</i> -type <i>orthacantha</i> ; dm: <i>Digera</i> -type <i>muricata</i> ; mono: Monocotyledonae; p: Poaceae; indet: indeterminable.	134

## Table des tableaux

<b>Table 2.1</b> – Regional vegetation, depositional environments, vegetation types and sub-types associated with the pollen samples present in each <i>k</i> -mean groups. 0 : van Breugel et al. (2015). Vegetation types and sub-types are based on 1: El-Moslimany (1983), 2: Friis et al. (2010), 3: Frey and Kürschner (1989), 4: <i>in situ</i> observations, 5: Barboni et al. (2019), 6: Bonnefille et al. (1987), 7: Mohammed et al. (2007), 8: Buchet (1982), and 9: Carr (1998).....	38
<b>Table 2.2</b> - Relationship between vegetation inferred by pollens and the characteristics of rivers. a : Personal observation using Google Earth images (1984 to 2020); 1 : Bonnefille et al. (1987); 2 : Buchet (1982); 3 : Carr (1998); 4 : Friis et al. (2010); 5 : Kobayashi et al. (2017). For abbreviations, see Table 2.1.....	54
<b>Table 2.3</b> - The 10 most characteristic pollen types of each vegetation type inferred by the pollen based of the T-values (Annex table 2.6). Taxa are listed from the most to the least important. For abbreviations, see Table 2.1.....	55
<b>Table 3.1</b> - Localisation, stratigraphy and results of V and V rivers scores. MMB assemblage: Maka'amitalu Basin assemblage ; KH-7: KadaHadar Member sub-member 7 ; KH-2: KadaHadar Member sub-member 2 ; KH-1: KadaHadar Member sub-member 1 ; DD-2: Denen Dora Member sub-member 2 ; DD-1: Denen Dora Member sub-member 1 ; SH-4: Sidi Hakoma sub-member 4 ; SH-3: Sidi Hakoma sub-member 3 ; SH-2: Sidi Hakoma sub-member 2 ; SH-1: Sidi Hakoma sub-member 1 ; UBM: Upper Basal Member ; BBM: Base of Basal Member. (1: Bonnefille et al., 1987, 2: Bonnefille et al., 2004, 3: Bonnefille, 2010, 4: López-Sáez and Domínguez-Rodrigo, 2009 and, 5: Saylor et al., 2019. For abbreviations of inferred paleovegetation, see Table 3.2.....	71
<b>Table 3.2</b> - Post-hoc vegetation with their abbreviation that can be identified with the V and V-river scores approach. 0: van Breugel et al. (2015), Micro-habitat types are based on 1: El-Moslimany (1983), 2: Friis et al. (2010), 3: Frey and Kürschner (1989), 4: <i>in situ</i> observations, 5: Barboni et al. (2019), 6: Bonnefille et al. (1987), 7: Mohammed et al. (2007), 8: Buchet (1982), 9: Carr (1998), and 10: Part 1 of this thesis.....	73

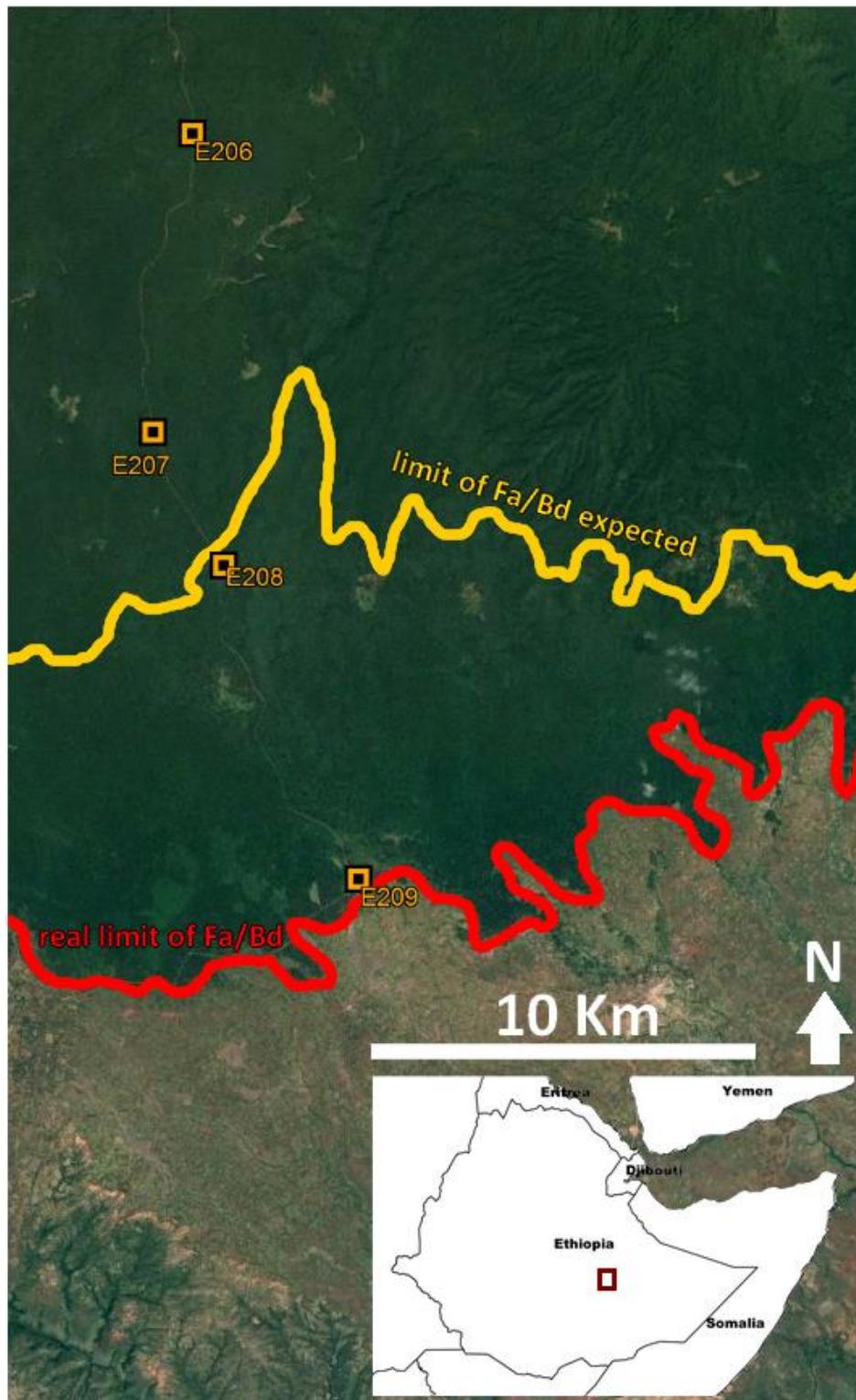
## Table des annexes

<b>Annexe figure 2.1</b> – Comparison between the limit of Fa/Bd expected by van Breugel et al. (2015) and the real limit observed with Google Earth.....	168
<b>Annexe figure 2.2</b> - Values of <i>V</i> classified in ascending order.....	169
<b>Annexe figure 2.3</b> - Comparison of K-means partitions for the number of abundance classes with the ssi criterion.....	170
<b>Annexe figure 2.4</b> - Comparison of K-means partitions for the number of pollen sample groups with the ssi criterion.....	170
<b>Annexe figure 2.5</b> – Comparison of hierarchical clustering with its version optimized by k-mean.....	171
<b>Annexe figure 2.6</b> – Silhouette plot represented by the ssi of each sample in each km-group for 21 km-groups.....	172
<b>Annexe figure 2.7</b> - Range of significant and non-significant score (cf., part 2.4.6.) for each river vegetation with a confidence interval of 95% represented by cyan and grey bars, respectively. The vegetation in parentheses indicate vegetation that have ranges of scores that overlap with range of significant score. For abbreviations, see Table 2.1.....	173
<b>Annexe figure 2.8</b> - Mean Decrease Accuracy for the 38 parameters used for reconstruct our 20 k-mean groups with Random Forest. For abbreviations, see Table 2.1..	174
<b>Annexe figure 2.9</b> - Precipitation seasonality (coefficient of variation) of the each identified vegetation based on their samples (WordClim2, 1970-2000, 0.93x0.93 km, Fick and Hijmans, 2017) with a confidence interval of 95%. For abbreviations, see Table 2.1.....	175
<b>Annexe figure 3.1</b> - <i>V</i> scores by vegetation for all samples. The blue bars represent the ranges of significant scores (95% confidence interval) and the grey bars represent the ranges of non-significant scores (95% confidence interval). The vegetation in brackets correspond to vegetation with <i>V</i> scores in the area of overlap of significant and non-significant values. The best score of each sample is circled in red. ....	176
<b>Annexe figure 3.2</b> – <i>V</i> -river scores by river vegetation for all samples. The blue bars represent the ranges of significant scores (95% confidence interval) and the grey bars represent the ranges of non-significant scores (95% confidence interval). The vegetation in brackets correspond to vegetation with <i>V</i> scores in the area of overlap of significant and non-significant values. The best score of each sample is circled in red.....	177
<b>Annexe figure 3.3</b> - Numerical version of stratigraphical section AV77 KM (= Hurda section) (south of KadadaMoumou Plateau, Ethiopia) with the possition of the Bonnefille's pollen samples by Jean-Jacques Tiercelin (22 January 1977), archives of the palynology team of CEREGE. ....	178
<b>Annexe figure 5.1</b> - Effect of transfer learning (the use of pretrained CNN) on the training with augmentation of CNN. Dataset: training pollen dataset; mapping:	

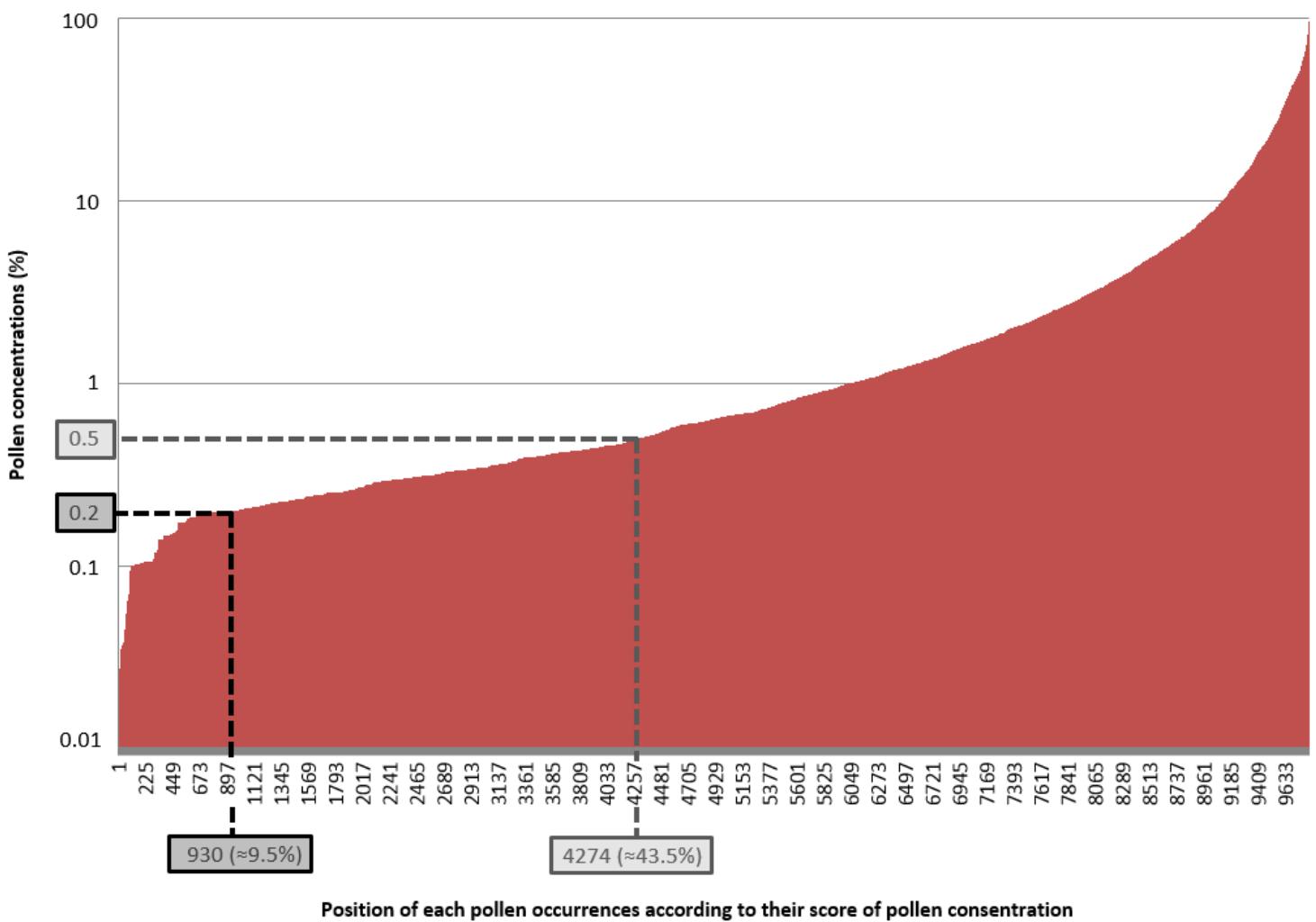
<i>Achyranthes</i> -type <i>aspera</i> , <i>Aerva</i> -type <i>javanica</i> , <i>Aerva</i> -type <i>lanata</i> , <i>Celosia</i> -type <i>argentea</i> , <i>Cyathula</i> -type <i>orthacantha</i> , <i>Digera</i> -type <i>muricata</i> , Cyperaceae, and Monocotyledonae.....	179
<b>Annexe figure 5.2</b> - Confusion matrix for the damaged pollen dataset (classes of step 2A (a) and classes of step 2B (b)) and the fossil pollen dataset (classes of step 2A (c) and classes of step 2B (d)). aa: <i>Achyranthes</i> -type <i>aspera</i> ; aj: <i>Aerva</i> -type <i>javanica</i> ; al: <i>Aerva</i> -type <i>lanata</i> ; amar: Amaranthaceae; c: Cyperaceae; ca: <i>Celosia</i> -type <i>argentea</i> ; co: <i>Cyathula</i> -type <i>orthacantha</i> ; dm: <i>Digera</i> -type <i>muricata</i> ; mono: Monocotyledonae; p: Poaceae; indet: indeterminable.....	180
<b>Annexe table 2.1</b> - Metadata of pollen samples (online). .....	181
<b>Annexe table 2.2</b> - Clustering results and vegetation information for each pollen sample (online).....	182
<b>Annexe table 2.3</b> - Climatic data, edaphic data, net primary production value, and % of tree cover for each pollen sample (online). .....	183
<b>Annexe table 2.4</b> - List of accepted pollen types and their synonyms.....	184
<b>Annexe table 2.5</b> - Raw pollen count with harmonized pollen types according to the Annex table 2.4 (online).....	189
<b>Annexe table 2.6</b> - list of Value Test (online). .....	189
<b>Annexe table 3.1</b> - Pollen counts used in this study without spores and undetermined pollen grains (online). .....	190
<b>Annexe table 3.2</b> - Age estimation and Age models (online) .....	190
<b>Annexe table 4.1</b> - Pollen counts and pollen type information on pollen habitus (A: trees or shrubs; AL: trees, lianas or climbing shrubs; I: undifferentiated; L: lianas; N: herbs; NL: climbing herbs; Nq aquatic herbs; PA: tree-like palms), water requirement (Hydro: hydrophyte; I: undifferentiated; Meso: mesophyte; Xero: xerophyte), salt tolerance (H: halophyte), and life cycle (An: annual; De: deciduous; Ev: evergreen; I: undifferentiated; Pe: perennial).....	190
<b>Annexe table 4.2</b> - Age values calculated with the age model based on $^{14}\text{C}$ dating.....	192
<b>Annexe table 4.3</b> - The Medieval Climate anomaly (MCA) and Little Ice Age (LIA) date ranges in Eastern Africa.....	192
<b>Annexe table 5.1</b> - Origin of pollen grains and distribution in the datasets. ka: thousand years. .....	193
<b>Annexe paper 1.1</b> - Springs, palm groves, and the record of early hominins in Africa.194	
<b>Annexe paper 1.2</b> - Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia .....	232

## Annexe des figures

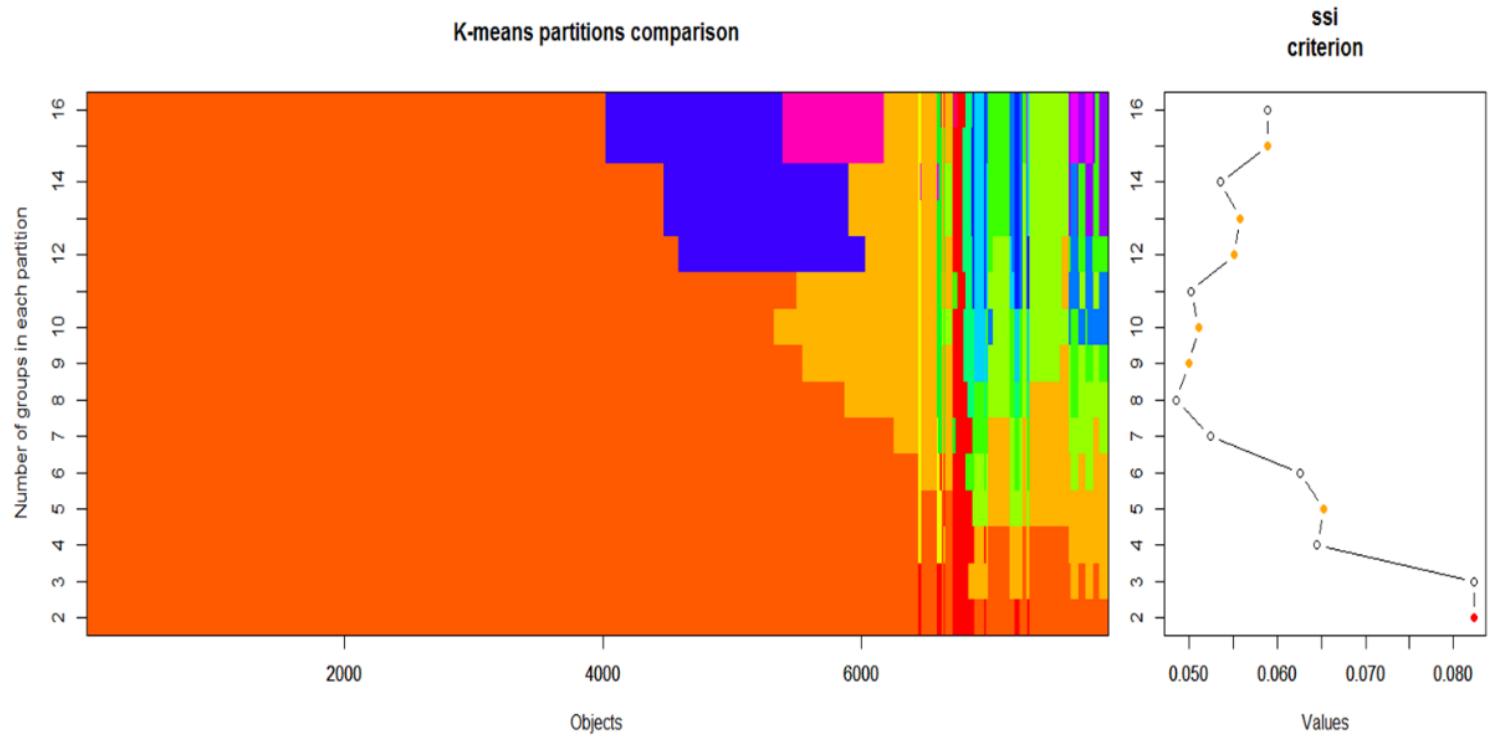
**Annexe figure 2.1** – Comparison between the limit of Fa/Bd expected by van Breugel et al. (2015) and the real limit observed with Google Earth.



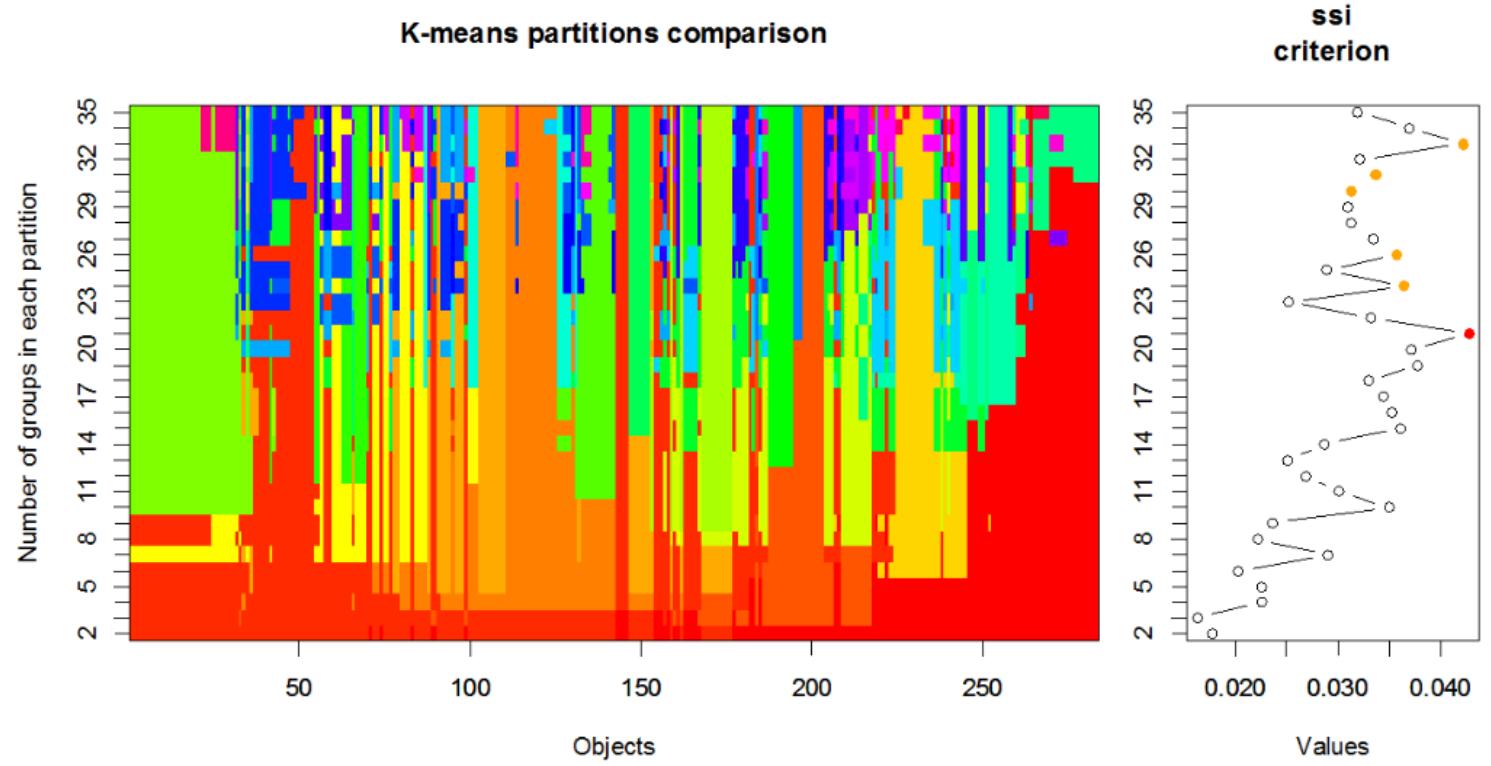
**Annexe figure 2.2 - Values of  $\vec{V}$  classified in ascending order.**



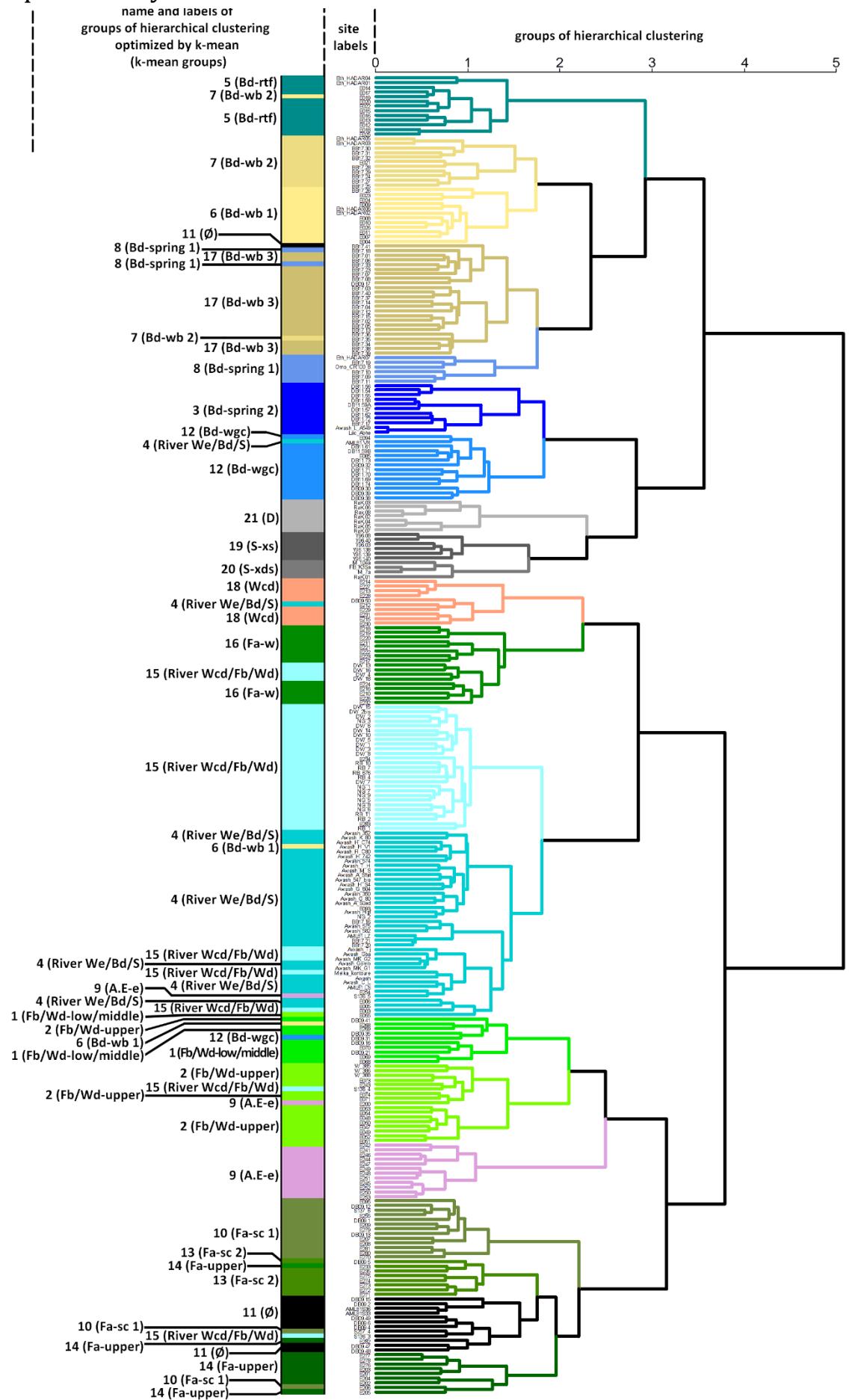
**Annexe figure 2.3** - Comparison of K-means partitions for the number of abundance classes with the ssi criterion.



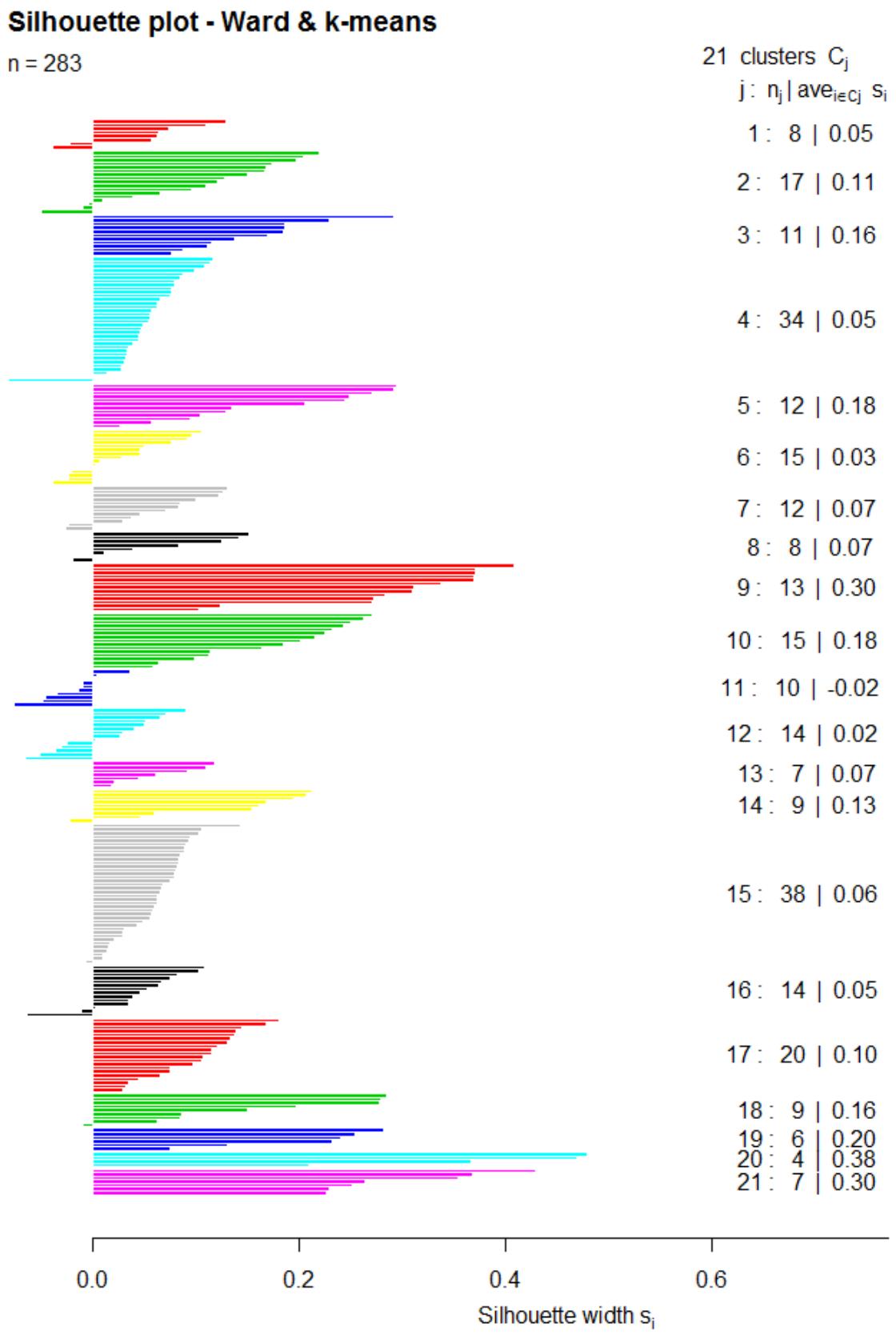
**Annexe figure 2.4** - Comparison of K-means partitions for the number of pollen sample groups with the ssi criterion.



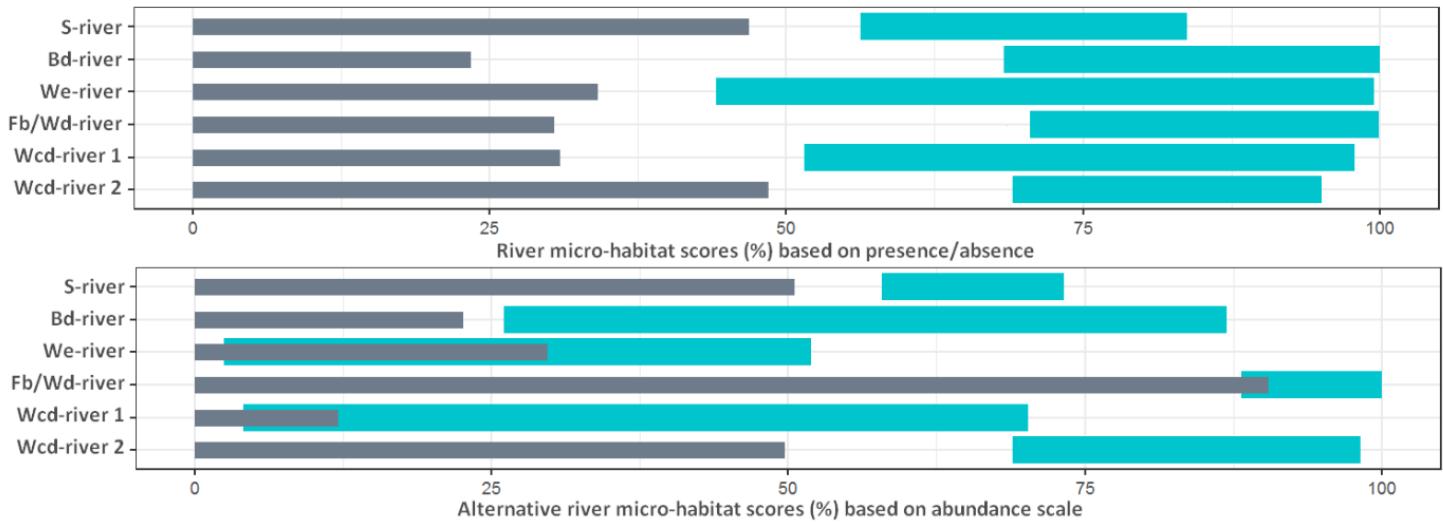
**Annexe figure 2.5 – Comparison of hierachical clustering with its version optimized by k-mean.**



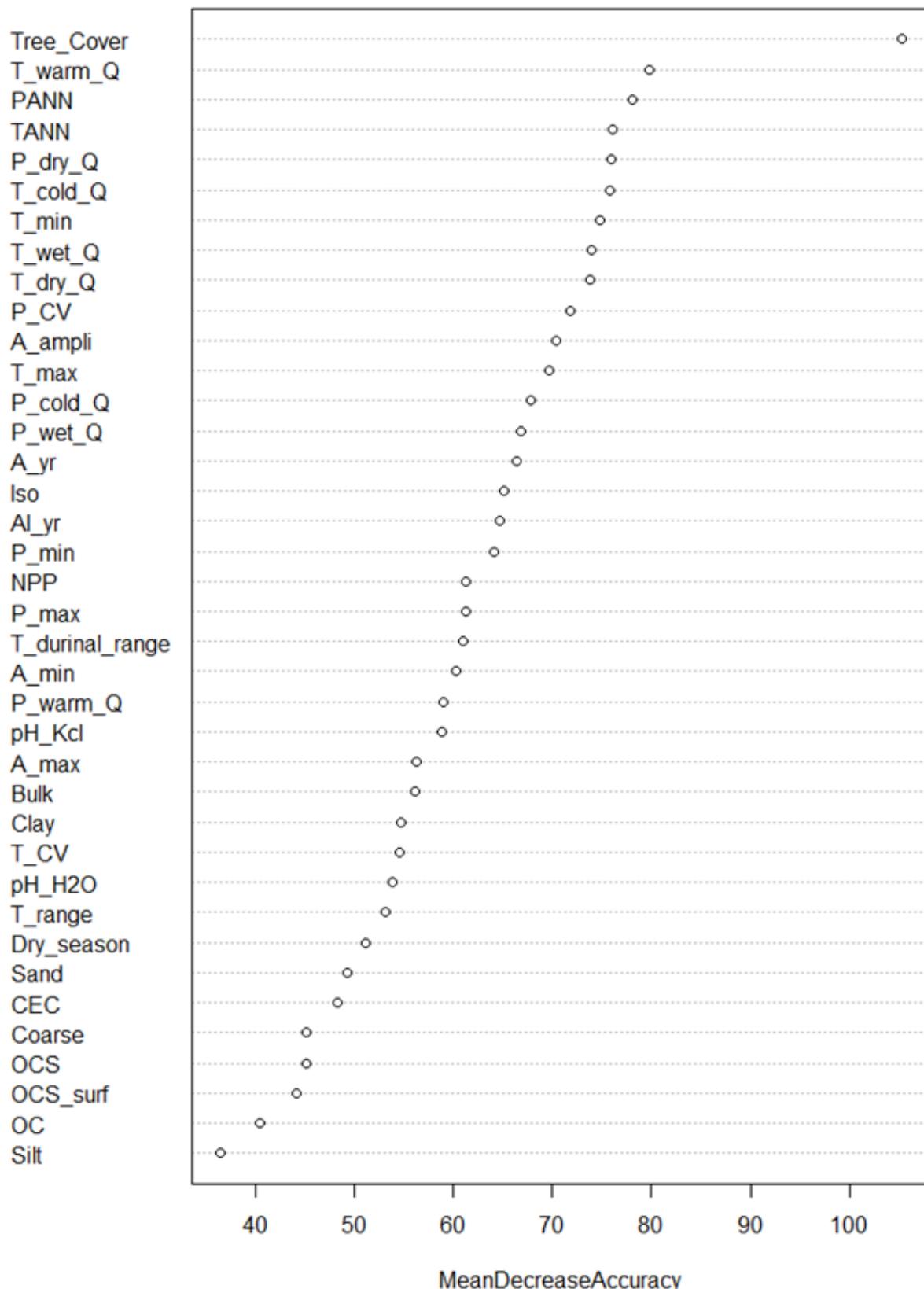
**Annexe figure 2.6** – Silhouette plot represented by the ssi of each sample in each km-group for 21 km-groups.



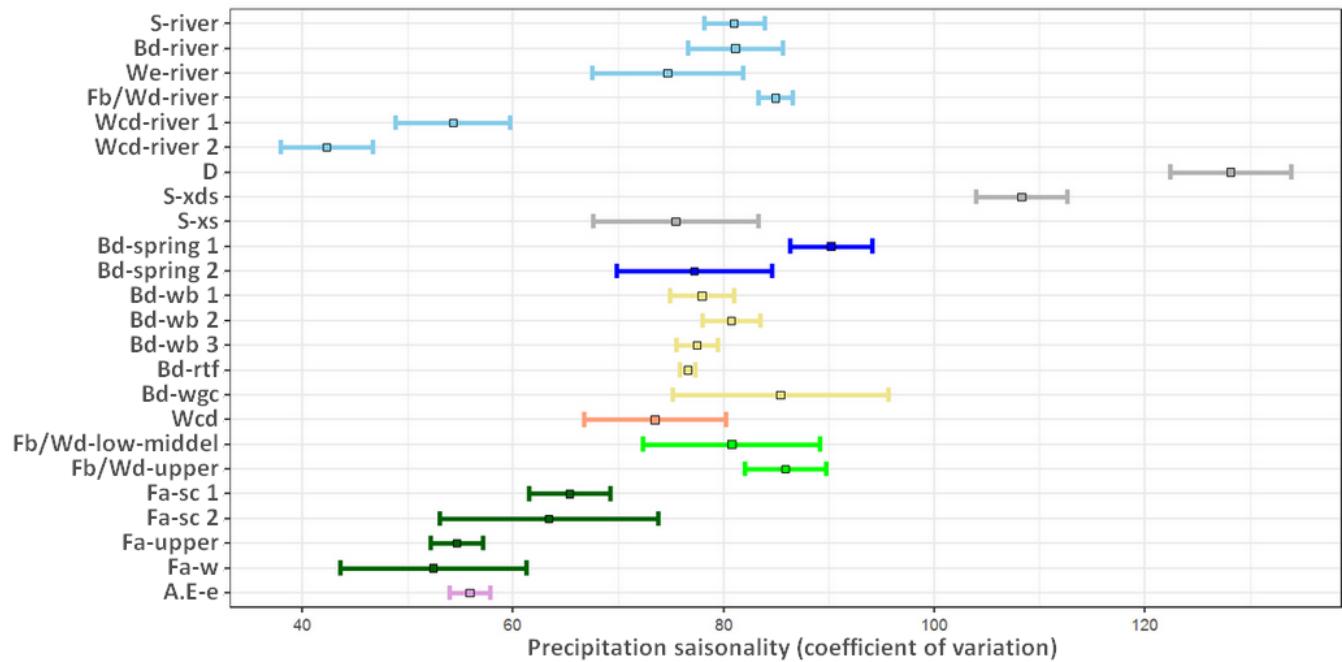
**Annexe figure 2.7** - Range of significant and non-significant score (cf., part 2.4.6.) for each river vegetation with a confidence interval of 95% represented by cyan and grey bars, respectively. The vegetation in parentheses indicate wegetation that have ranges of scores that overlap with range of significant score. For abbreviations, see Table 2.1.



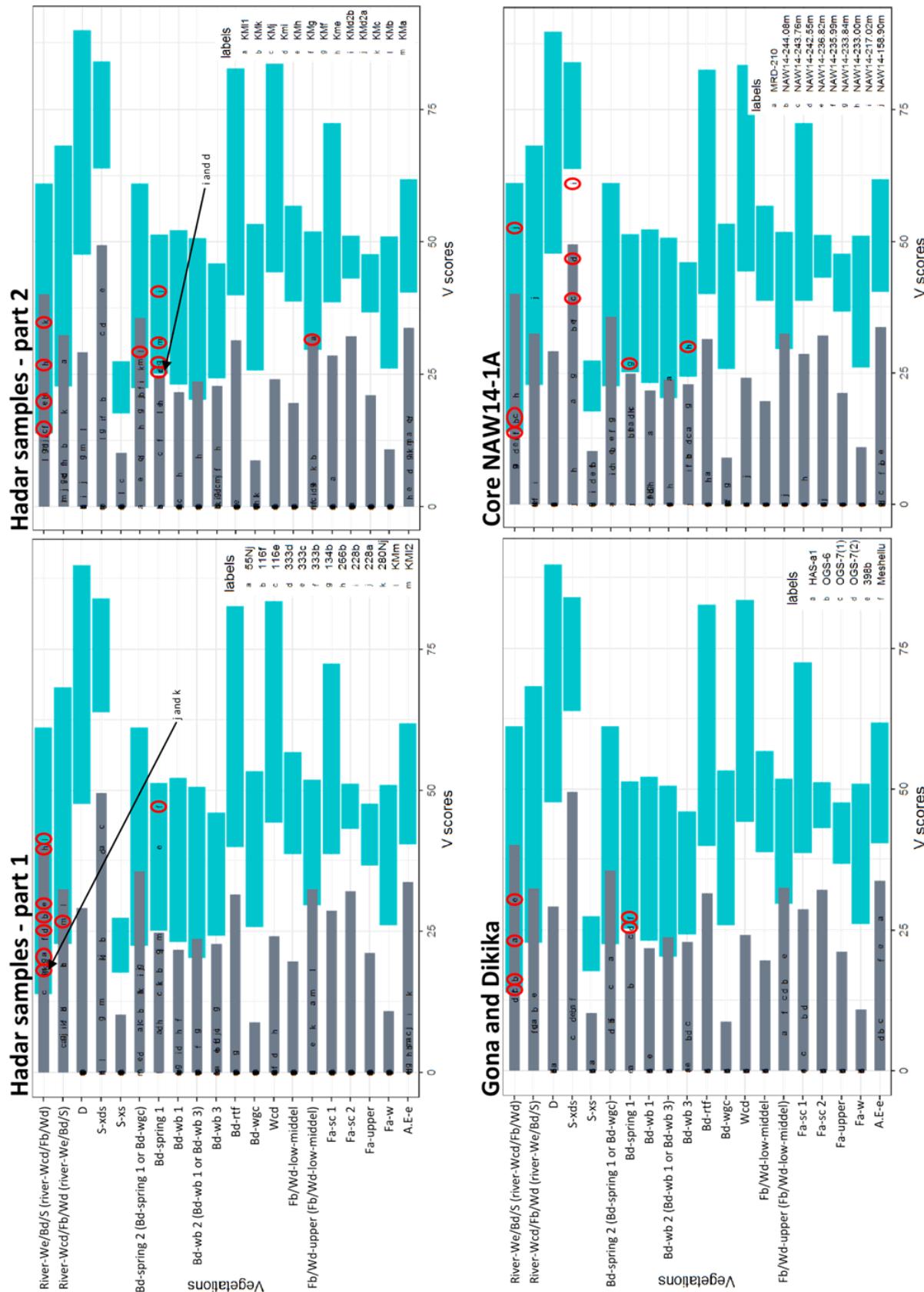
**Annexe figure 2.8** - Mean Decrease Accuracy for the 38 parameters used for reconstruct our 20 k-mean groups with Random Forest. For abbreviations, see Table 2.1.



**Annexe figure 2.9** - Precipitation seasonality (coefficient of variation) of the each identified vegetation based on their samples (WordClim2, 1970-2000, 0.93x0.93 km, Fick and Hijmans, 2017) with a confidence interval of 95%. For abbreviations, see Table 2.1.

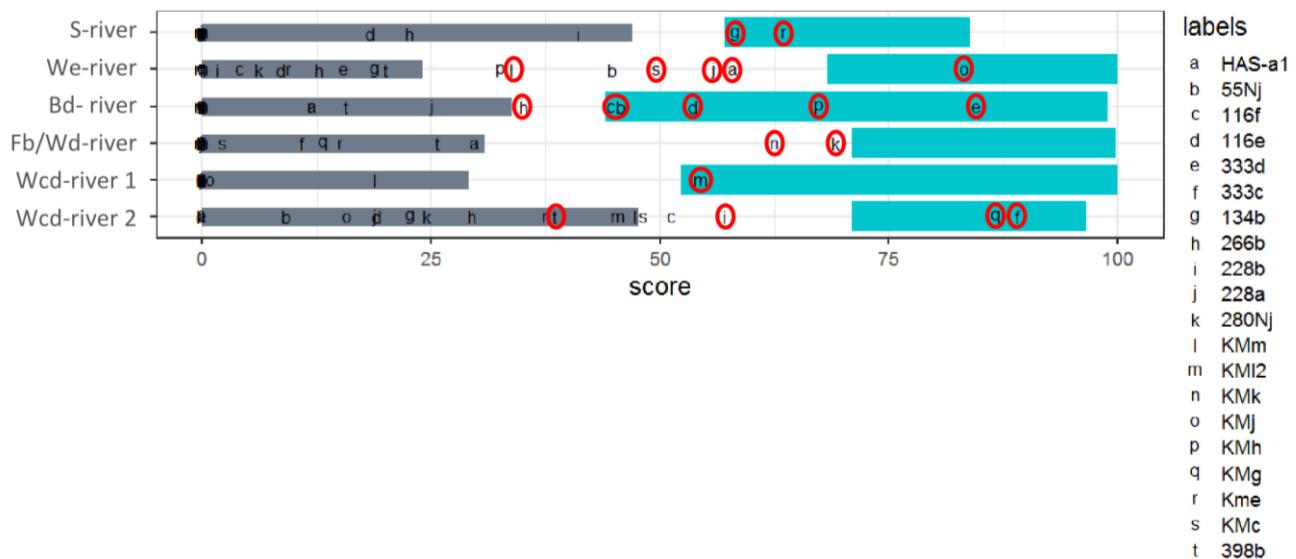


**Annexe figure 3.1** - V scores by vegetation for all samples. The blue bars represent the ranges of significant scores (95% confidence interval) and the grey bars represent the ranges of non-significant scores (95% confidence interval). The vegetation in brackets correspond to vegetation with V scores in the area of overlap of significant and non-significant values. The best score of each sample is circled in red.

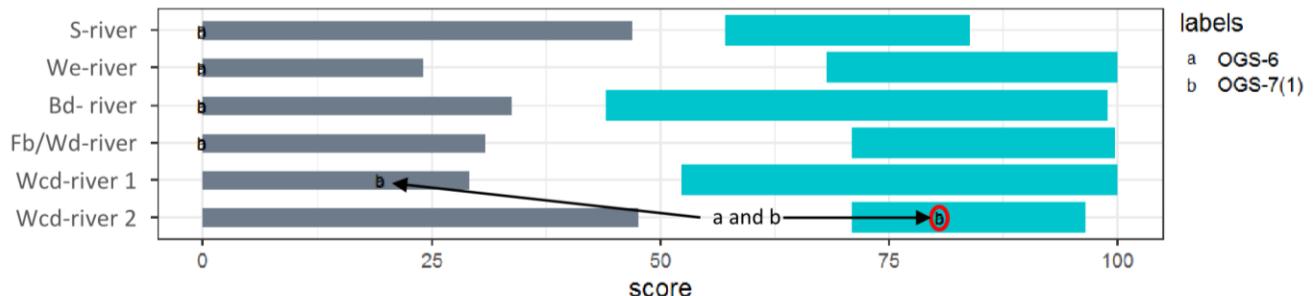


**Annexe figure 3.2 – V-river scores by river vegetation for all samples.** The blue bars represent the ranges of significant scores (95% confidence interval) and the grey bars represent the ranges of non-significant scores (95% confidence interval). The vegetation in brackets correspond to vegetation with V scores in the area of overlap of significant and non-significant values. The best score of each sample is circled in red.

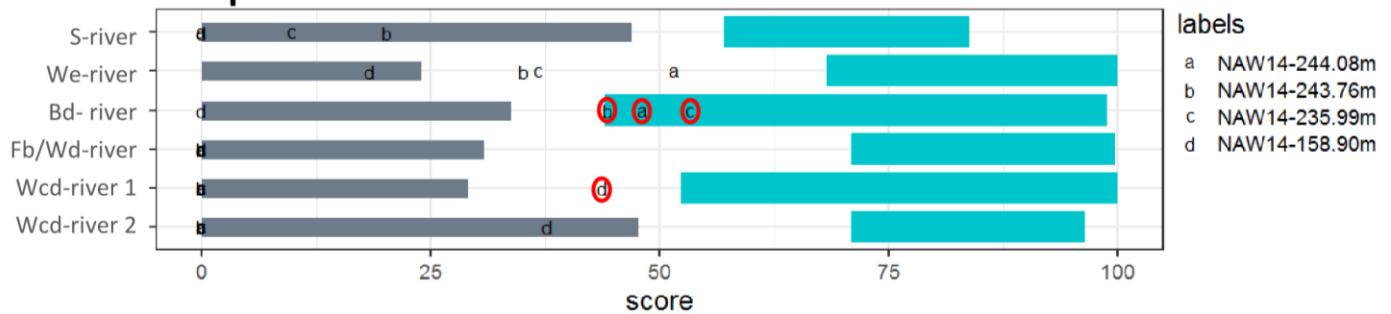
### Hadar samples



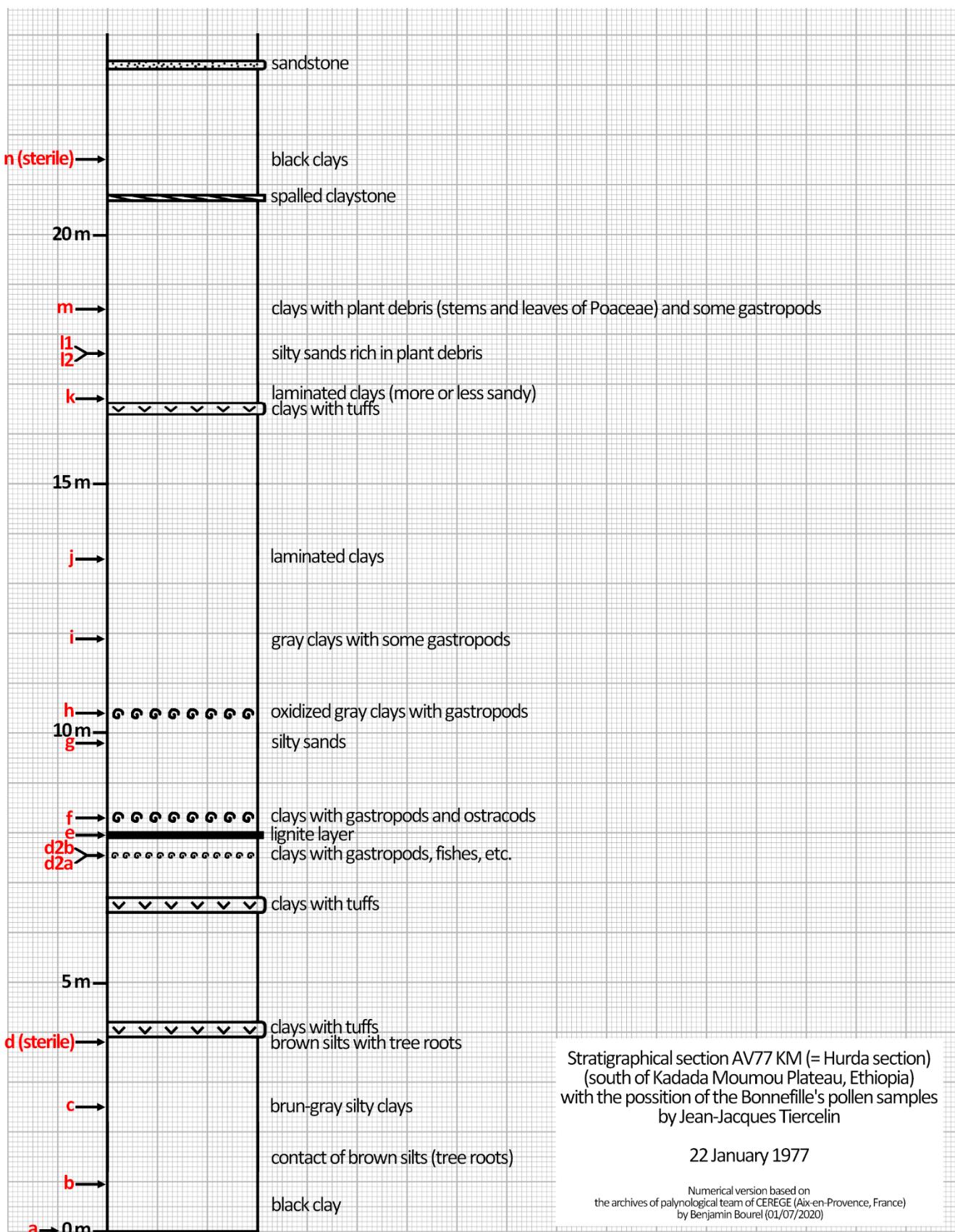
### Gona samples



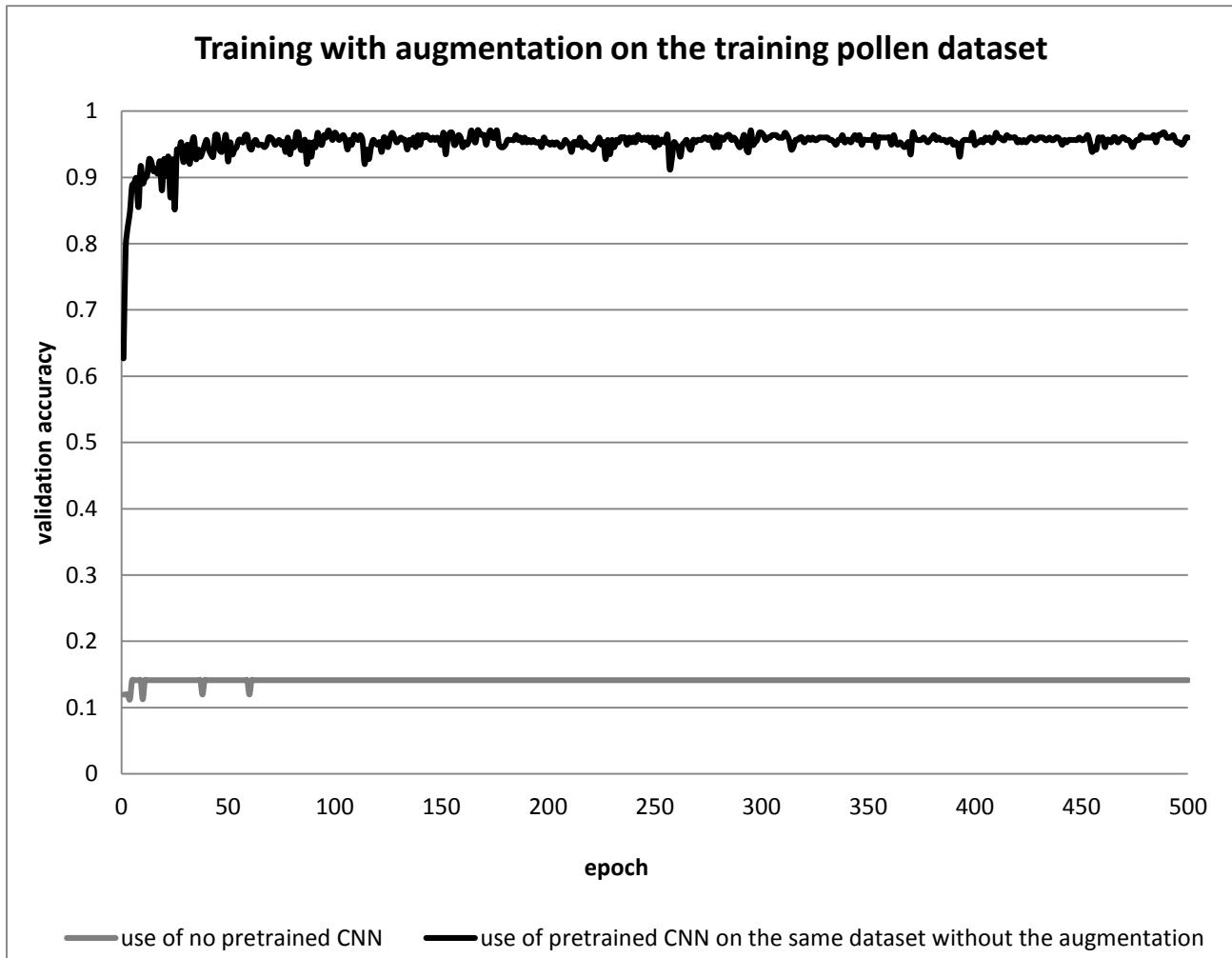
### Samples of core NAW14-1A



**Annexe figure 3.3** - Numerical version of stratigraphical section AV77 KM (= Hurda section) (south of KadadaMoumou Plateau, Ethiopia) with the position of the Bonnefille's pollen samples by Jean-Jacques Tiercelin (22 January 1977), archives of the palynology team of CEREGE.

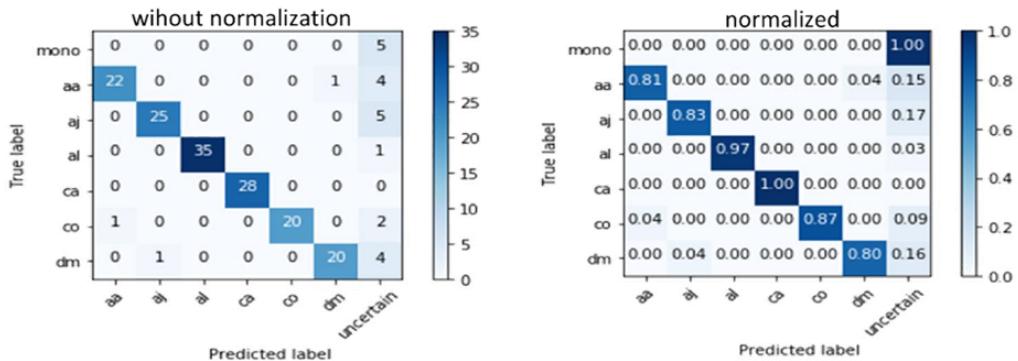


**Annexe figure 5.1** - Effect of transfer learning (the use of pretrained CNN) on the training with augmentation of CNN. Dataset: training pollen dataset; mapping: *Achyranthes-type aspera*, *Aerva-typejavanica*, *Aerva-typelanata*, *Celosia-type argentea*, *Cyathula-type orthacantha*, *Digera-type muricata*, Cyperaceae, and Monocotyledonae.

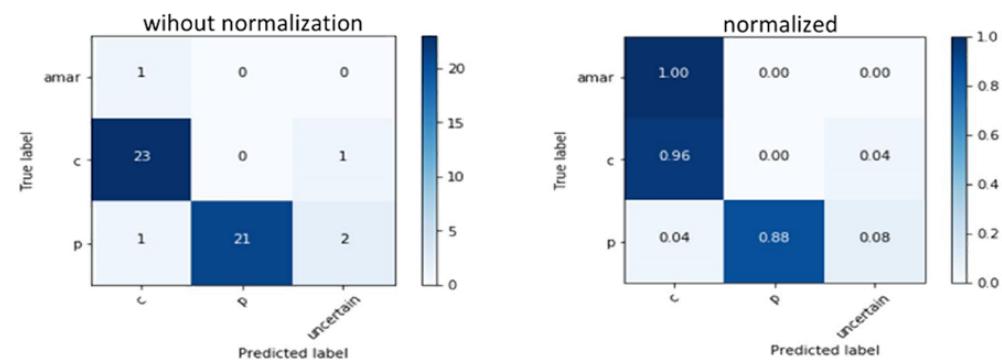


**Annexe figure 5.2** - Confusion matrix for the damaged pollen dataset (classes of step 2A (a) and classes of step 2B (b)) and the fossil pollen dataset (classes of step 2A (c) and classes of step 2B (d)). aa: *Achyranthes*-type *aspera*; aj: *Aerva*-type *javanica*; al: *Aerva*-type *lanata*; amar: Amaranthaceae; c: Cyperaceae; ca: *Celosia*-type *argentea*; co: *Cyathula*-type *orthacantha*; dm: *Digera*-type *muricata*; mono: Monocotyledonae; p: Poaceae; indet: indeterminable.

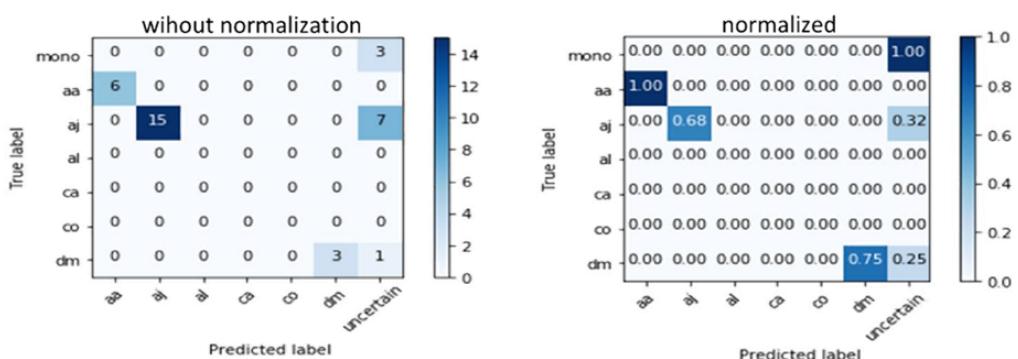
a. Classes of step 3A - damaged pollen



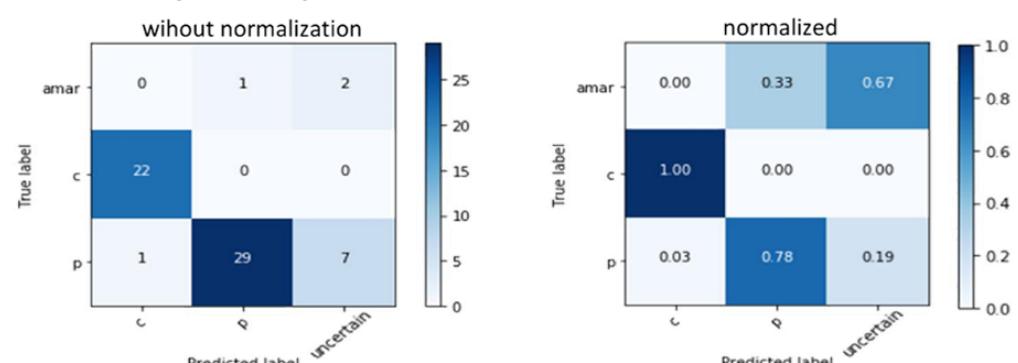
b. Classes of step 3B - damaged pollen



c. Classes of step 3A - fossil pollen dataset



d. Classes of step 3B - fossil pollen dataset



## Annexe des tableaux

**Nota bene :** Certain tableaux étant beaucoup très volumineux (jusqu'à 300 colonnes et 300 linges), il n'a pas été possible de les intégré de façon acceptable à ce mémoire de thèse. Lorsque c'est le cas, un lien de téléchargement du fichier avec le tableau sera proposé.

### Annexe table 2.1 - Metadata of pollen samples (online).

**Column 1:** Pollen sample number.

**Column 2:** Sample label.

**Column 3:** Type of sample (surface soil or mud collected in river; old: from the APD or the literature; new: this study).

**Column 4:** Sampling locality.

**Column 5:** Elevation at the sampling site (in m).

**Column 6:** Latitude at the sampling site (in DD).

**Column 7:** Longitude at the sampling site (in DD).

**Column 8:** Correction of the sample altitude (in m). This corresponds to the difference between the altitude given in column 5 and the altitude archived in the APD or indicated in the original publication in column 18.

**Column 9:** Correction of sample location (in km). This corresponds to the difference between the position given in columns 6-7 and the position archived in the APD or indicated in the original publication in column 18.

**Column 10:** Accuracy of location data (in m) obtained by taking into account the accuracy of GPS data for the "new" samples and the accuracies given by the authors for "old" samples

**Column 11:** Year when sampling was carried out; may be approximate for the "old" samples.

**Column 12:** Database where pollen counts are available.

**Column 13:** APD folder that contains the pollen counts (when counts are present).

**Column 14:** Sample label in the APD database (when counts are present).

**Column 15:** Sample label in the CEREGE database (when counting sheets are present).

**Column 16:** pollen analyst.

**Column 17:** Original label of the sample, as it appears in original publication (cf., column 18).

**Column 18:** Main references where the sample is positioned and/or described. These are the short references, the full references are in the "References" section.

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 2.2** - Clustering results and vegetation information for each pollen sample (online).

**Column 1:** Pollen sample number.

**Column 2:** Sample label.

**Column 3:** Elevation

**Column 4:** Width of the silhouette. The greater the width of the silhouette of a sample, the more appropriate the group membership, and a negative silhouette width may indicate a misclassification.

**Column 5:** Number of the k-means group of the sample obtained after hierarchical clustering optimized by k-mean.

**Column 6:** Neighboring k-mean group of samples.

**Column 7:** Number of the k-means group accepted after comparison with the column 5 and 6

**Column 8:** Type of sample (surface soil or mud collected in river; old: from the APP or the literature; new: this study).

**Column 9:** Sampling locality.

**Column 10:** Notes of the people who sampled and publication informations (Barboni et al., 2019 ; Bonnefille, 1972 ; Bonnefille and Buchet, 1986 ; Bonnefille et al. 1987, 1993, 2004 ; Bonnefille and Riollet, 1988 ; El-Moslimany 1983 ; Gallotti et al., 2013 ; Lézine, 1981 ; Lézine et al, 1998 ; Mohammed, 1992 ; Mohammed and Bonnefille, 2002 ; Mohammed et al., 2007).

**Column 11:** Potential vegetation based on van Breugel et al. (2015) and Friis et al (2010) for African samples, and Frey and Kürschner (1989) for Arabian samples (n: 252 to 268). For abbreviations, see "Notes".

**Column 12:** Correction of potential vegetation based on satellite images of Google Earth. For abbreviations, see "Notes".

**Column 13:** Important edaphic informations for the sample (fed by spring = the spring is not on site but at 250-500 m, the site is fed by short rivers or surface runoffs) (Ayenew et al., 2008 ; Barboni et al., 2019 ; Bonnefille, 1972 ; Bonnefille and Buchet, 1986 ; Bonnefille et al. 1987, 1993, 2004 ; Bonnefille and Riollet, 1988 ; Cuthbert et al., 2017 ; El-Moslimany 1983 ; Frey and Kürschner, 1989 ; Le Turdu et al., 1999 ; Lézine, 1981 ; Lézine et al, 1998 ; Mohammed, 1992 ; Mohammed and Bonnefille, 2002 ; Mohammed et al., 2007).

**Column 14:** Changes since sample collection. Based on comparisons of satellite images with the "time travel" function of Google Earth .

**Column 15:** Accepted observed vegetation type for each sample.

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 2.3** - Climatic data, edaphic data, net primary production value, and % of tree cover for each pollen sample (online).

**Column 1:** Pollen sample number.

**Column 2:** Sample label.

**Column 3:** Number of the k-means group accepted.

**Column 4:** Post-hoc inference of vegetation types and sub-types based on km-groups. For the nomenclature, see Table 2.1.

**Column 5:** indicates if the sample is classified as "river" in the initial k-means clustering analysis.

**Column 6:** Latitude at the sampling site (in DD).

**Column 7:** Longitude at the sampling site (in DD).

**Column 8:** Type of sample (surface soil or mud collected in river)

**Column 9:** Elevation at the sampling site (in m).

**Columns 10-19:** Temperature data ( $^{\circ}\text{C}$ ) of WordClim 2.0 ( $0.93 \times 0.93$  km). Mean temperature of coldest quarter (T\_cold\_Q), mean temperature of warmest quarter (T\_warm\_Q), mean temperature of wettest quarter (T\_wet\_Q), annual temperature range (T\_range), min. temperature of coldest month (T\_min), max. temperature of warmest month (T\_max), temperature seasonality or coefficient of variation (T\_CV), isothermality (Iso), annual mean diurnal range (T\_diurnal\_range), annual mean temperature (TANN).

**Columns 20-28:** Precipitation data (mm/month) of WordClim 2.0 ( $0.93 \times 0.93$  km). Mean precipitation of coldest quarter (P\_cold\_Q), mean precipitation of warmest quarter (P\_warm\_Q), mean precipitation of wettest quarter (P\_wet\_Q), precipitation seasonality or coefficient of variation (P\_CV), precipitation of wettest month (P\_min), precipitation of driest month (P\_max), annual mean precipitation (PANN).

**Columns 29-40:** Precipitation data for each month (mm/month) of WordClim 2.0 ( $0.93 \times 0.93$  km). Prec\_XX is the mean precipitation for the month number XX of the year.

**Column 41:** Length of the dry season, number of months per year with less than 30 mm of precipitation. Based on WordClim 2.0 ( $0.93 \times 0.93$  km).

**Columns 42-46:** AET/PET in equilibrium (alpha coefficient) of Global High-Resolution Soil-Water Balance ( $0.93 \times 0.93$  km) and aridity index data of Global High-Resolution Soil-Water Balance ( $0.93 \times 0.93$  km). Mean annual AET/PET (alpha\_yr), AET/PET of driest month (alpha\_min), AET/PET of wettest month (alpha\_max), difference between alpha\_max and alpha\_min (alpha\_ampli), aridity index of year (AI\_yr).

**Column 47:** Percentage of tree cover, data of Global Map ( $0.46 \times 0.46$  km).

**Column 48:** Net primary production (kg of C/m<sup>2</sup>/yr) data of MODIS17A3 ( $0.93 \times 0.93$  km).

**Columns 49-53:** Physical soil properties at 15 cm of Soilgrids ( $0.25 \times 0.25$  km). Bulk density in km/cm<sup>3</sup> (Bulk), % of clay (Clay), % of coarse fragment (Coarse), % of silt (Silt), % of sand (Sand).

**Columns 54-59:** Chemical soil properties at 15 cm except column 56 (OCS : between 15-30cm) and column 57 (OCS\_surf : 0cm) of Soilgrids ( $0.25 \times 0.25$  km). Cation exchange capacity in cmolc/kg (CEC), organic carbon in g/kg (OC), organic carbon stock in T/ha

(OCS), surface organic carbon stock in T/ha (OCS\_surf), pH in H<sub>2</sub>O (pH\_H2O), pH in KCl (pH\_KCl).

**Column 60:** Soil types (WRB classification) based on Soilgrids (0.25×0.25 km). Haplic Alisols (Acri), Aluandic Andosols (Alu\_Ando), Aric Regosols (Ari\_Rego), Calcaric Regosols (Cal\_Rego), Haplic Calcisols (Calci), Haplic Cambisols (Cambi), Haplic Ferralsols (Ferral), Haplic Fluvisols (Fluvi), Haplic Solonchaks (Halo), Haplic Leptosols (Lepto), Lithic Leptosols (Lit\_Lepto), Haplic Luvisols (Luvi), Haplic Nitisos-Rhodic(Niti), Protic Arenosols (Pro\_Aerno), Rendzic Leptosols (Ren\_Lepto), Haplic Vertisols (Verti).

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

## Annexe table 2.4 - List of accepted pollen types and their synonyms.

**Column 1:** Family of pollen-type accepted.

**Column 2:** Pollen types accepted following Vincens et al. (2007).

**Column 3:** Original pollen types as given by the palynologists (synonymous or obsolete pollen-types).

**Notes :** The pollen types were harmonized between the APD samples, the non-APD samples and our new samples. Synonymous pollen types were grouped together such as Cleome-type brachycarpa and Dipterygium-type glaucum. Pollen types difficult to differentiate in the modern samples and rarely differentiated in the fossils register were also grouped together such as Nuxia/Dobera. We have limited the taxonomic precision of some families such as Asteraceae. Indeed, in the case of Asteraceae, 1/3 samples were identified only at the subfamily level: Cichorieae or Tubuliflorae. The spores were excluded from our analyses because some authors did not count them and for others authors, it is difficult to know if they did not count or if they were not present.

1	2	3
family	pollen_types_accepted	original_pollen_types
ACANTHACEAE	Acanthus-type	Blepharis-type
ACANTHACEAE	Duosperma-type	Duosperma
ACANTHACEAE	Dyschoriste-type radicans	Dyschoriste-type
ACANTHACEAE	Hygrophila-type	Hygrophila, Hygrophila abyssinica-type
ACANTHACEAE	Hypoestes-type	Hypoestes/Peristrophe
ACANTHACEAE	Justicia striata-type	Justicia-type striata
ACANTHACEAE	Justicia-type flava	Justicia_caerulea
ACANTHACEAE	Megalochlamis revoluta-type	Ecbolium
ACANTHACEAE	Mellera-type lobulata	Mellera-type
ACANTHACEAE	Mimulopsis-type solmsii	Mimulopsis-type
ACANTHACEAE	Phaulopsis-type imbricata	Phaulopsis-type
ACANTHACEAE	Ruellia-type patula	Ruellia

<b>ACANTHACEAE</b>	Ruttya-type fruticosa	Ruttya-type
<b>AIZOACEAE</b>	Aizoaceae undiff.	Corbicichonia decumbens/Trianthema
<b>AMARANTHACEAE</b>	Alternanthera	Alternanthera sessilis-type
<b>AMARANTHACEAE</b>	Celosia-type trigyna	Celosia-type schweinfurthiana
<b>AMARANTHACEAE</b>	Cyathula-type orthacantha	Sericocomopsis-type, Sericocomopsis-type pallida
<b>AMARANTHACEAE</b>	Psilotrichum-type elliottii	Psilotrichum-type
<b>AMARANTHACEAE</b>	Pupalia-type lappacea	Pupalia-type
<b>AMARANTHACEAE/CHE NOPODIACEAE/CARYO PHYLLACEAE</b>	Amaranthaceae/Chenopodiacea e/Caryophyllaceae undiff.	Amaranthaceae/Chenopodiaceae undiff., Amaranthaceae/Caryophyllaceae undiff., Amaranthaceae undiff.
<b>ANACARDIACEAE</b>	Ozoroa-type insignis	Ozoroa-type
<b>ANACARDIACEAE</b>	Rhus-type vulgaris	Rhus-type natalensis
<b>ANNONACEAE</b>	Annona-type senegalensis	Annona senegalensis-type
<b>APIACEAE (UMBELLIFERA)</b>	Hydrocotyle mannii-type	Hydrocotyle, Hydrocotyle hirta-type, Hydrocotyle ranunculoides-type
<b>APIACEAE (UMBELLIFERA)</b>	Steganotaenia-type araliacea	Steganotaenia-type
<b>APIACEAE (UMBELLIFERA)</b>	Steganotaenia-type araliacea	Heteromorpha-type
<b>APOCYNACEAE</b>	Acokanthera-type schimperi	Acokanthera-type
<b>APOCYNACEAE</b>	Adenium-type obesum	Adenium
<b>APOCYNACEAE</b>	Oncinotis-type	Oncinotis-type nitida
<b>APOCYNACEAE</b>	Strophanthus-type mirabilis	Funtumia-type africana
<b>ARALIACEAE</b>	Cussonia-type	Cussonia, Schefflera lukwanguensis-type
<b>ASTERACEAE</b>	Cichorieae	Crepis-type, Crepis/Launaea, Hirpicium-type diffusum, Vernonia schimperi, Vernonia perrottetii, Vernonieae undiff.
<b>ASTERACEAE</b>	Tubuliflorae	Asteraceae, Aaronsohnia-type pubescens, Ambrosia-type, Artemisia, Carduus, Carthamus-type, Crassocephalum/Senecio, Echinops, Filago, Guizotia-type abyssinica, Pulicaria, Rhanterium, Sphaeranthes-type ukembensis, Stoebe kilimandscharica-type, Tarchonanthus_camphoratus, Xanthium-type
<b>BIGNONIACEAE</b>	Kigelia-type	Kigelia africana
<b>BORAGINACEAE</b>	Cordia	Cordia_africana, Cordia_sinensis
<b>BORAGINACEAE</b>	Echium-type	Echium plantagineum-type
<b>BORAGINACEAE</b>	Ehretia cymosa-type	Ehretia
<b>BORAGINACEAE</b>	Heliotropium steudneri-type	Heliotropium, Heliotropium indicum-type, Heliotropium ovalifolium-type, Heliotropium sudanicum-type
<b>BORAGINACEAE</b>	Lithospermum-type afromontanum	Arnebia hispidissima-type, Lithospermum afromontanum
<b>BRASSICACEAE</b>	Brassicaceae undiff.	Arabis, CRUCIFERA, Farsetia stenoptera, Schouwia purpurea
<b>BUXACEAE</b>	Buxus	Buxus hildebrandtii-type
<b>BURSERACEAE</b>	Commiphora	Commiphora africana-type, Commiphora edulis-type, Commiphora madagascarensis-type
<b>CAESALPINIACEAE</b>	Caesalpinia-type trothae	Caesalpinia-type

<b>CAESALPINIACEAE</b>	Cassia-type	Cassia-type didymobotrya, Cassia-type italica, Cassia-type mimosoides, Cassia-type occidentalis
<b>CAESALPINIACEAE</b>	Hymenostegia-type pellegrinii	Tamarindus-type indica
<b>CAESALPINIACEAE</b>	Parkinsonia-type	Parkinsonia aculeata-type
<b>CAESALPINIACEAE</b>	Piliostigma	Bauhinia rufescens-type, Piliostigma thonningii-type
<b>CAPPARIDACEAE</b>	Boscia-type	Cadaba-type glandulosa, Cadaba-type rotundifolia
<b>CAPPARIDACEAE</b>	Capparis	Capparis cartilaginea-type, Capparis fascicularis-type, Capparis tomentosa-type
<b>CAPPARIDACEAE</b>	Cleome-type	Cleome-type gynandra
<b>CAPPARIDACEAE</b>	Cleome-type brachycarpa	Dipterygium-type glaucum
<b>CAPPARIDACEAE</b>	Crateva-type adansonii	Crateva adansonii
<b>CAPPARIDACEAE</b>	Maerua-type crassifolia	Maerua/Ritchiea, Maerua-type oblongifolia, Thilachium-type
<b>CARYOPHYLLACEAE</b>	Drymaria cordata-type	Drymaria
<b>CARYOPHYLLACEAE</b>	Paronychia	Paronychia-type
<b>CARYOPHYLLACEAE</b>	Polycarpaea-type	Polycarpon
<b>CELASTRACEAE</b>	Maytenus senegalensis-type	Maytenus
<b>CELASTRACEAE/HIPPO CRATEACEAE</b>	Celastraceae/Hippocrateaceae undiff.	Celastraceae undiff., Hippocrateaceae undiff.
<b>CHENOPODIACEAE</b>	Suaeda monoica-type	Suaeda
<b>CISTACEAE</b>	Helianthemum	Helianthemum baciferum-type
<b>CLUSIACEAE</b>	Garcinia granulata*	Garcinia (only for counts of Bonnefille's team)
<b>COMBRETACEAE</b>	Anogeissus-type leiocarpa	Anogeissus-type leiocarpus
<b>COMBRETACEAE</b>	Combretum-type molle	Combretum/Terminalia, Combretum-type, Terminalia-type
<b>CONVOLVULACEAE</b>	Hildebrandtia-type obcordata	Cressa-type cretica, Cuscuta-type
<b>CUCURBITACEAE</b>	Cucumis dipsaceus-type	Cucumis
<b>DRACAENACEAE</b>	Dracaena	Dracaena schizantha-type, Dracaena steudneri-type
<b>EPHEDRACEAE</b>	Ephedra	Ephedra alata-type, Ephedra distachya-type, Ephedra fragilis-type
<b>ERICACEAE</b>	Ericaceae undiff.	Erica arborea/Agauria salicifolia
<b>EUPHORBIACEAE</b>	Acalypha	Acalypha indica-type, Acalypha psilostachya-type
<b>EUPHORBIACEAE</b>	Andrachne aspera-type	Andrachne
<b>EUPHORBIACEAE</b>	Bridelia-type	Bridelia ferruginea-type
<b>EUPHORBIACEAE</b>	Clutia	Clutia abyssinica-type
<b>EUPHORBIACEAE</b>	Croton-type	Jatropha-type
<b>EUPHORBIACEAE</b>	Euphorbia tirucalli-type	Euphorbia-type tirucalli
<b>EUPHORBIACEAE</b>	Euphorbiaceae undiff.	Euphorbia-type, Phyllanthus-type
<b>EUPHORBIACEAE</b>	Euphorbia-type indica	Euphorbia hypericifolia-type
<b>EUPHORBIACEAE</b>	Macaranga-type	Macaranga-type capensis
<b>EUPHORBIACEAE</b>	Phyllanthus-type boehmii	Phyllanthus_amarus, Phyllanthus-type fischeri, Phyllanthus-type fraternus
<b>FABACEAE</b>	Fabaceae undiff.	PAPILIONOIDEAE

<b>FABACEAE</b>	Calpurnia aurea	Calpurnia-type aurea
<b>FABACEAE</b>	Indigofera schimperi-type	Indigofera
<b>FABACEAE</b>	Millettia-type ferruginea	Millettia-type
<b>FABACEAE</b>	Rhynchosia-type malacophylla	Rhynchosia-type
<b>FABACEAE</b>	Sesbania sesban	Sesbania
<b>FABACEAE</b>	Trifolium-type acaule	Trifolium-type, Trifolium-type burchellianum
<b>FABACEAE</b>	Vicia	Lathyrus/Vicia, Vicia faba, Vigna luteola
<b>GENTIANACEAE</b>	Gentianaceae undiff.	Swertia-type
<b>GENTIANACEAE</b>	Swertia usambarensis-type	Swertia abyssinica-type
<b>HYMENOCARDIACEAE</b>	Hymenocardia	Hymenocardia acida-type
<b>HYPERICACEAE</b>	Hypericum	Hypericum annulatum-type, Hypericum peplidifolium-type, Hypericum quartinianum-type
<b>LAMIACEAE (LABIATAE)</b>	Basilicum-type polystachyon	Basilicum polystachyon/Hoslundia opposita, Satureja-type kilimandschari
<b>LAMIACEAE (LABIATAE)</b>	Hyptis pectinata-type	Hyptis
<b>LAMIACEAE (LABIATAE)</b>	Lamiaceae	Lamiaceae_III, Lamiaceae_VI
<b>LINACEAE</b>	Linum	Linum strictum-type
<b>LOGANIACEAE</b>	Buddleja	Buddleja polystachya-type
<b>LOGANIACEAE/SALVADORACEAE</b>	Nuxia/Dobera	Dobera-type, Dobera-type glabra, Dobera/Salvadora, Nuxia-type, Nuxia-type congesta, Salvadoria, Salvadoria/Nuxia, Salvadoria persica-type, Salvadoraceae undiff.
<b>LORANTHACEAE</b>	Plicosepalus-type	Plicosepalus-type sagittifolius
<b>LORANTHACEAE</b>	Tapinanthus-type globiferus	Tapinanthus-type
<b>MALVACEAE</b>	Abutilon-type	Abutilon, Gossypium
<b>MALVACEAE</b>	Hibiscus/Pavonia	Hibiscus, Hibiscus-type micranthus, Hibiscus micranthus-type, Pavonia
<b>MELIACEAE</b>	Ekebergia-type capensis	Ekebergia-type
<b>MELIANTHACEAE</b>	Bersama abyssinica-type	Bersama
<b>MENISPERMACEAE</b>	Cissampelos_mucronata	Cissampelos-type
<b>MENISPERMACEAE</b>	Cocculus-type	Cocculus, Cocculus hirsutus-type, Cocculus pendulus-type
<b>MENISPERMACEAE</b>	Stephania-type abyssinica	Stephania-type
<b>MENISPERMACEAE</b>	Tiliacora-type funifera	Tiliacora-type
<b>MIMOSACEAE</b>	Acacia	Acacia, Acacia I-type, Acacia III-type
<b>MIMOSACEAE</b>	Dichrostachys cinerea-type	Dichrostachys
<b>MORACEAE</b>	Dorstenia-type	Dorstenia kameruniana-type, Dorstenia-type scaphigera
<b>MORACEAE</b>	Milicia-type excelsa	Antiaris-type toxicaria, Morus-type mesozygia
<b>MYRICACEAE</b>	Myrica	Myrica humilis-type
<b>MYRSINACEAE</b>	Maesa lanceolata-type	Maesa
<b>MYRSINACEAE</b>	Rapanea melanophloeos-type	Rapanea
<b>MYRTACEAE</b>	Eucalyptus-type	Eucalyptus-type globulus
<b>MYRTACEAE</b>	Eugenia/Syzygium	Myrtaceae undiff., Eugenia-type, Syzygium-type, Syzygium-type guineense
<b>OLEACEAE</b>	Chionanthus mildbraedii-type	Chionanthus niloticus-type

<b>OLEACEAE</b>	Olea europaea-type	Olea europaea-type ssp africana, Ligustrum lucidum-type
<b>PALMAE</b>	Borassus-type	Borassus-type aethiopum
<b>PALMAE</b>	Hyphaene-type	Hyphaene-type, Hyphaene-type thebaica
<b>PALMAE</b>	Phoenix reclinata-type	Phoenix, Phoenix dactylifera-type
<b>PITTOSPORACEAE</b>	Pittosporum abyssinicum-type	Pittosporum
<b>PLANTAGINACEAE</b>	Plantago	Plantago africana-type, Plantago ciliata-type, Plantago coronopus-type, Plantago lanceolata-type
<b>POACEAE</b>	Poaceae undiff.	Sinarundinaria alpina, Secale cereale
<b>POLYGONACEAE</b>	Polygonum senegalense-type	Polygonum pulchrum-type
<b>POLYGONACEAE</b>	Rumex	Calligonum polygonoides, Rumex vesicarius-type
<b>PORTULACACEAE</b>	Portulacaceae undiff.	Portulaca-type
<b>POTAMOGETONACEAE</b>	Potamogeton	Potamogeton perctinatus-type
<b>PRIMULACEAE</b>	Anagallis angustiloba-type	Anagallis
<b>PRIMULACEAE</b>	Primulaceae undiff.	Lysimachia, Primula verticillata
<b>PROTEACEAE</b>	Protea-type	Proteaceae undiff.
<b>RANUNCULACEAE</b>	Ranunculus multifidus-type	Ranunculus stagnalis-type
<b>RESEDACEAE</b>	Resedaceae undiff.	Caylusea-type abyssinica
<b>RHAMNACEAE</b>	Rhamnaceae undiff.	Ziziphus-type
<b>ROSACEAE</b>	Prunus africana-type	Prunus
<b>ROSACEAE</b>	Rubus pinnatus-type	Rubus
<b>RUBIACEAE</b>	Anthospermum	Anthospermum herbaceum-type
<b>RUBIACEAE</b>	Hymenodictyon-type floribundum	Hymenodictyon_floribundum
<b>RUBIACEAE</b>	Kohautia caespitosa-type	Kohautia
<b>RUBIACEAE</b>	Pavetta abyssinica-type	Pavetta
<b>RUBIACEAE</b>	Rubia-type cordifolia	Rubia-type, Galium-type
<b>RUBIACEAE</b>	Tarenna-type graveolens	Tarenna-type
<b>RUBIACEAE</b>	Vangueria-type madagascariensis	Vangueria-type
<b>RUTACEAE</b>	Teclea-type	Vepris-type, Vepris-type dainellii, Vepris-type nobilis, Zanthoxylum-type chalybeum
<b>RUTACEAE</b>	Zanthoxylum/Fagaropsis	Fagaropsis-type angolensis, Zanthoxylum-type, Zanthoxylum-type usambarensis
<b>SAPINDACEAE</b>	Allophylus abyssinicus-type	Allophylus
<b>SAPINDACEAE</b>	Blighia unijugata-type	Blighia
<b>SAPINDACEAE</b>	Dodonaea viscosa-type	Dodonaea
<b>SAPINDACEAE</b>	Zantha	Zantha golungensis-type
<b>SAPOTACEAE</b>	Pouteria-type	Aphania-type senegalensis, Pouteria-type adolfi-friedericii
<b>SAPOTACEAE</b>	Manilkara-type butugii	Manilkara-type
<b>SIMAROUBACEAE</b>	Brucea	Brucea antidyserterica-type
<b>SOLANACEAE</b>	Lycium-type	Datura, Lycium
<b>SOLANACEAE</b>	Solanum-type	Capsicum, Solanum-type giganteum
<b>STERCULIACEAE</b>	Dombeya-type	Dombeya-type burgessiae
<b>THYMELAEACEAE</b>	Gnidia-type chrysantha	Gnidia-type
<b>TILIACEAE</b>	Corchorus-type	Corchorus-type fascicularis
<b>TILIACEAE</b>	Grewia-type	Grewia
<b>TILIACEAE</b>	Tiliaceae undiff.	Triumphetta-type

<b>ULMACEAE</b>	Celtis	Celtis africana-type, Holoptelea grandis
<b>URTICACEAE</b>	Laportea-type aestuans	Laportea-type
<b>URTICACEAE</b>	Pilea-type bambuseti	Pilea-type
<b>URTICACEAE</b>	Urticaceae D-type (D-type= pollen diporate)	Pilea-type bambuseti
<b>URTICACEAE</b>	Urticaceae P-type (P-type= pollen periporate)	Laportea-type aestuans
<b>URTICACEAE</b>	Urticaceae Q-type (Q-type= pollen quadriporate)	Urera-type hypselodendron, Urtica-type, Urticaceae undiff.
<b>URTICACEAE</b>	Urticaceae T-type (T-type= pollen triporate)	Debregeasia-type saeneb, Droguetia, Pouzolzia-type
<b>VERBENACEAE</b>	Lantana-type ukambensis	Phyla-type, Phyla-type nodiflora
<b>VERBENACEAE</b>	Lantana-type viburnoides	Lantana-type, Lantana-type ukambensis, Lippia-type
<b>VITACEAE</b>	Cissus quadrangularis-type	Cissus
<b>ZYGOPHYLLACEAE</b>	Zygophyllum simplex-type	Zygophyllum

**Annexe table 2.5** - Raw pollen count with harmonized pollen types according to the Annexe table 2.4 (online).

**Column 1:** The new pollen types comparatively to Vincens et al. (2007).

**Column 2:** Total number of pollen grain counted in the data set (Column 10-291).

**Column 3:** Number of occurrences in the data set (Column 10-291).

**Column 4:** Plant habitus. Tree (T), shrub (S), herb (H), liana (L), aquatic (a), parasitic (p) (after Vincens et al. 2007).

**Column 5:** Pollen habitus. Tree and shrub (A), tree, shrub, liana (AL), liana (L), herb (N), climbing herb (NL), aquatic herb (Nq), undifferentiated (I), palm liana (PL), palm tree (PA), palm undifferentiated (PI), parasitic (p) . Based on Vincens et al. (2007).

**Column 6:** Family of pollen type.

**Column 7:** Pollen types accepted (Vincens et al. 2007).

**Column 8:** Label of pollen type used in the analyses.

**Columns 9-291:** raw pollen count in samples.

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 2.6** - list of Value Test (online).

**Column 1:** Label of pollen type used in the analyses.

**Columns 2-22:** Value Test (Lebart et al., 1995) of pollen type for Km-group # and Post-hoc inference of vegetation (abundance scale)

**Columns 23-28:** Value Test (Lebart et al., 1995) of pollen type for River groups (presence/absence)

**Columns 29-34:** Value Test (Lebart et al., 1995) of pollen type for River groups (abundance scale)

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 3.1** - Pollen counts used in this study without spores and undetermined pollen grains (online).

**Column 1:** pollen types identified in the samples of our study that have no equivalence in the 249 modern samples used to calibrate the approach of V and V-river scores in Part 1 of this thesis.

**Column 2:** Family of pollen-type accepted.

**Column 3:** Pollen types accepted in Part 1 of this thesis following Vincens et al. (2007).

**Columns 4-45:** raw pollen count in samples with the samples ordered from the most recent to the oldest.

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 3.2** - Age estimation and Age models (online)

**Annexe table 3.2.a** - Age estimation and relative stratigraphic position of the pollen samples

**Annexe table 3.2.b** - Age model for Hurda samples

**Annexe table 3.2.c** - Age model for NAW14-1A samples

Like : <https://github.com/Beniofh/online-annexe-of-thesis>

**Annexe table 4.1** - Pollen counts and pollen type information on pollen habitus (A: trees or shrubs; AL: trees, lianas or climbing shrubs; I: undifferentiated; L: lianas; N: herbs; NL: climbing herbs; Nq: aquatic herbs; PA: tree-like palms), water requirement (Hydro: hydrophyte; I: undifferentiated; Meso: mesophyte; Xero: xerophyte), salt tolerance (H: halophyte), and life cycle (An: annual; De: deciduous; Ev: evergreen; I: undifferentiated; Pe: perennial).



**Annexe table 4.2 - Age values calculated with the age model based on  $^{14}\text{C}$  dating.**

Section of core KN1A	Laboratory	Laboratory sample no.	$^{13}\text{C}/^{12}\text{C} \text{‰}$	Percent Modern Carbon (pMC)	Conventional Radiocarbon Age (BP)	Calibrated Radiocarbon Age (cal AD)	
						IntCal 13	IntCal 20
22-23 cm	Beta Analytic (Miami, Florida, USA)	Beta 463027	-21.5	93.50±0.35	540±30	1413±24	1414±23
30-31 cm	Beta Analytic (Miami, Florida, USA)	Beta 463028	-21.3	91.31±0.34	730±30	1261±37	1238±9
37.5-38 cm	Artemis instrument, UMS 2572, Laboratoire de mesure du carbone 14 (Saclay, France)	SacA 33589	-26.0	87.72±0.29	1055±30	984±40	991±42
42-43 cm	Artemis instrument, UMS 2572, Laboratoire de mesure du carbone 14 (Saclay, France)	SacA 33590	-22.8	86.21±0.83	1190±80	830±158	832±161

**Annexe table 4.3 - The Medieval Climate anomaly (MCA) and Little Ice Age (LIA) date ranges in Eastern Africa.**

	Southern Hemisphere	
	MCA	Main phase of LIA
Russell and Johnson, 2007	1000 to 1200	1400 to 1750
Haug et al., 2001		1350 to 1740
Russell et al., 2007	975 to 1200	1500 to 1700-1750
Stager et al., 2005		1550 to 1750
Wolff et al. 2011	900-950 to 1200	1300-1400 to 1750
<i>Core period</i>	<i>950-1200</i>	<i>1350-1400 to 1700-1750</i>

**Annexe table 5.1** - Origin of pollen grains and distribution in the datasets.  
ka: thousand years.

family	pollen type	origin	grains	dataset
Amaranthaceae	<i>Achyranthes* aspera</i>	slide n° 4302 of the reference pollen collection of CEREGE	130	training pollen
Amaranthaceae	<i>Achyranthes* aspera</i>	slide n° 4302 of the reference pollen collection of CEREGE	33	intact pollen
Amaranthaceae	<i>Achyranthes* aspera</i>	slide n° 865 of the reference pollen collection of CEREGE	27	damaged pollen
Amaranthaceae	<i>Achyranthes* aspera</i>	core BH33 - Urmia - Iran - depth 5 cm ( 6 ka)	6	fossil pollen
Amaranthaceae	<i>Aerva* javanica</i>	slide n° 2813 of the reference pollen collection of CEREGE	134	training pollen
Amaranthaceae	<i>Aerva* javanica</i>	slide n° 2813 of the reference pollen collection of CEREGE	34	intact pollen
Amaranthaceae	<i>Aerva* javanica</i>	slide n° 3419 of the reference pollen collection of CEREGE	31	damaged pollen
Amaranthaceae	<i>Aerva* javanica</i>	sample MRD-210 - Godaya Valley - Woronso-Mille - Ethiopia (Pliocene)	25	fossil pollen
Amaranthaceae	<i>Aerva* lanata</i>	slide n° 2618 of the reference pollen collection of CEREGE	138	training pollen
Amaranthaceae	<i>Aerva* lanata</i>	slide n° 2618 of the reference pollen collection of CEREGE	35	intact pollen
Amaranthaceae	<i>Aerva* lanata</i>	slide n° 4263 of the reference pollen collection of CEREGE	36	damaged pollen
Amaranthaceae	<i>Celosia* argentea</i>	slide n° 5055 of the reference pollen collection of CEREGE	139	training pollen
Amaranthaceae	<i>Celosia* argentea</i>	slide n° 5055 of the reference pollen collection of CEREGE	35	intact pollen
Amaranthaceae	<i>Celosia* argentea</i>	slide n° 143 of the reference pollen collection of CEREGE	28	damaged pollen
Amaranthaceae	<i>Cyathula* orthacantha</i>	slide n° 5060 of the reference pollen collection of CEREGE	139	training pollen
Amaranthaceae	<i>Cyathula* orthacantha</i>	slide n° 5060 of the reference pollen collection of CEREGE	35	intact pollen
Amaranthaceae	<i>Cyathula* orthacantha</i>	slide n° 2435 of the reference pollen collection of CEREGE	25	damaged pollen
Amaranthaceae	<i>Digera* muricata</i>	slide n° 3551 of the reference pollen collection of CEREGE	146	training pollen
Amaranthaceae	<i>Digera* muricata</i>	slide n° 3551 of the reference pollen collection of CEREGE	36	intact pollen
Amaranthaceae	<i>Digera* muricata</i>	slide n° 2180 of the reference pollen collection of CEREGE	25	damaged pollen
Amaranthaceae	<i>Digera* muricata</i>	core BH33 - Urmia - Iran - depth 38 to 50 cm ( 30-35 ka)	4	fossil pollen
Cyperaceae	Cyperaceae	slide n° 2559 of the reference pollen collection of CEREGE	138	training pollen
Cyperaceae	Cyperaceae	slide n° 2559 of the reference pollen collection of CEREGE	34	intact pollen
Cyperaceae	Cyperaceae	slide n° 1572 of the reference pollen collection of CEREGE	25	damaged pollen
Cyperaceae	Cyperaceae	core BH33 - Urmia - Iran - depth 5 cm (6 ka)	25	fossil pollen
Poaceae	Poaceae	slide n° 2362 of the reference pollen collection of CEREGE	138	training pollen
Poaceae	Poaceae	slide n° 2362 of the reference pollen collection of CEREGE	34	intact pollen
Poaceae	Poaceae	slide n° 2382 of the reference pollen collection of CEREGE	28	damaged pollen
Poaceae	Poaceae	sample MRD-210 - Godaya Valley - Woronso-Mille - Ethiopia (Pliocene)	37	fossil pollen

## Annexe des publications

### Annexe paper 1.1 - Springs, palm groves, and the record of early hominins in Africa



### Springs, palm groves, and the record of early hominins in Africa

Doris Barboni, Gail Ashley, Benjamin Bourel, Hector Arraiz, Jean-Charles Mazur

#### ► To cite this version:

Doris Barboni, Gail Ashley, Benjamin Bourel, Hector Arraiz, Jean-Charles Mazur. Springs, palm groves, and the record of early hominins in Africa. *Review of Palaeobotany and Palynology*, Elsevier, 2019, 266, pp.23-41. 10.1016/j.revpalbo.2019.03.004 . hal-02099545

HAL Id: hal-02099545

<https://hal.archives-ouvertes.fr/hal-02099545>

Submitted on 24 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted manuscript now published in Review of Palaeobotany and Palynology Volume 266, July 2019, Pages 23-41 (2019) – Special Issue on “Plants and Human Evolution”

## Springs, palm groves, and the record of early hominins in Africa

Doris Barboni<sup>1\*</sup>, Gail M. Ashley<sup>2</sup>, Benjamin Bourel<sup>1</sup>, Hector Arraiz<sup>1,3</sup>, Jean-Charles Mazur<sup>1</sup>

<sup>1</sup> CEREGE, Aix-Marseille University, CNRS, IRD, Coll. France, INRA, Technopole Arbois-Méditerranée, BP80, 13545 Aix en Provence cedex 4, France

<sup>2</sup> Earth and Planetary Sciences, Rutgers University, 610 Taylor Rd., Piscataway, NJ 08854-8066, USA

<sup>3</sup> Department of Prehistory, Complutense University of Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

\* Corresponding author: +334 42 97 17 66; Email address: barboni@cerege.fr (D. Barboni)

### Highlights

- Groundwater systems locally create wet and wooded microhabitats in arid Eastern Africa.
- But, how important are groundwater-fed microhabitats to the early hominin record?
- Geological context, vegetation, and microfossils from modern springs were analyzed.
- Fossil springs were found associated with hominin remains in >50 different localities.
- Groundwater systems play essential roles at the species and ecosystem levels.

### Abstract

Hominins evolved in Africa during a period of overall regional cooling, drying and increasingly variable climate. Despite prevailing regional aridity since the mid-Miocene, data show that early hominins *Sahelanthropus tchadensis*, *Orrorin tugenensis*, and *Ardipithecus ramidus* lived in environments made of mosaics of grasslands, mixed grasslands, woodlands, and forests, where wooded habitats were maintained by edaphic rather than regional (climatic) humidity. Groundwater systems (springs, seeps, shallow aquifers) and surface water (rivers, lakes), locally create wetter and more wooded environments in addition to that supported by precipitation alone. However, edaphically sustained woodlands are rare to missing in most published paleoecological interpretations of hominin archeological sites. To explore the importance of groundwater to the record of hominins in Africa, we provide newly acquired field data from spring sites in the Awash Valley, Ethiopia, and Lake Eyasi-Lake Manyara region, Tanzania, and re-evaluate published data from the *Ardipithecus*-bearing Aramis Member, Ethiopia.

Results show that 1) in arid Eastern Africa, a wide variety of microhabitats such as groundwater-fed wetlands, *Hyphaene* palm woodlands, *Phoenix reclinata* palm woodlands, and structurally complex and species-rich forest patches exist due to local variability of geologic, topographic and hydrologic conditions. 2) These microhabitats carry some characteristic pollen and phytolith signals, that may be easily masked by the signal of surrounding grass-dominated shrublands and grasslands. 3) The Aramis Member (Awash Valley, Ethiopia), which is to date, the best documented paleo-groundwater ecosystem, is not a riparian habitat. It is one of > 50 examples (within 22 geographically distinct areas) in Africa and the Middle East where evidence of groundwater systems co-exist with hominin and/or archeological remains. Springs are commonly localized features of limited area within a landscape, but provide ecological continuity through time and diverse microhabitats, some of which may be densely forested.

At the local scale, springs create microclimates, distinctive vegetation, and increase soil nutrients, species richness, structural complexity, and provide habitat for animals. At the landscape scale, they represent hydro-refugia favoring increased connectivity among animals and allowing migrations during dry periods. We conclude that in the East African Rift where low, highly seasonal rainfall and high evaporative demand limit vegetation growth in many areas, groundwater-fed zones create diverse microhabitats and play a major role in ecosystem functioning. It is likely that, within a context of increasing aridity and expansion of grass-dominated open habitats during the Mio-Pliocene, early hominins and many other animals viewed edaphically sustained woodlands as attractive habitats.

**Keywords:** Hominin; Paleoenvironment; *Ardipithecus*; Pliocene; Groundwater; Hydro-refugia

## 1. Introduction

The hominin phylogenetic tree is often put in parallel with records characterizing the global climate changes, such as the record of  $\delta^{18}\text{O}$  a proxy for global ocean temperatures and polar ice volume (e.g. DeMenocal, 1995; Potts, 2013), or records of  $\delta^{13}\text{C}$  in paleosols as a proxy for C<sub>4</sub> plant biomass abundance (e.g. Feakins and deMenocal, 2010; Levin, 2015; Quade et al., 1989). Climate above all environmental factors determines the spatial distribution of vegetation types at the continental and regional scales (e.g. Box, 1981; Pearson and Dawson, 2003), which subsequently determines the type of plant resources available to consumers as well as the spatial and temporal availability of those resources. It is thus hypothesized that large climatically driven vegetation changes, such as the spread of open habitats and the expansion of C<sub>4</sub> grass-dominated biomes were responsible for the Late Miocene mammalian evolution, including that of our lineage (see review on the origin of the savanna hypothesis by Bender et al., 2012, and references cited therein). In East Africa the expansion of C<sub>4</sub> grasses is approximately dated at 10 Ma (Feakins et al., 2013; Uno et al., 2016). This profound change in vegetation composition and structure at the continental scale likely induced a shift in the diet of Proboscideans which started to include C<sub>4</sub> plants in their (browsing) diet as early as 9.9 Ma, and became grazers at 7 Ma (Uno et al., 2016). Similarly, environmental change could have led some Primates to adopt terrestrial bipedalism, a purportedly cost effective locomotion in open environments (e.g. Sockol et al., 2007; Steudel-Numbers and Tilkens, 2004). This so-called savanna hypothesis that had germinated in our collective consciousness

since Lamarck published the first ideas in 1809 (Bender et al., 2012; Lamarck, 1809), has been challenged by some, as contextual data associated with Late Miocene and Early Pliocene hominins *Sahelanthropus tchadensis*, *Orrorin tugenensis*, and *Ardipithecus ramidus* indicate wooded and forested, rather than open environments (Pickford and Senut, 2001; Vignaud et al., 2002; White et al., 2009b). The savanna hypothesis is further challenged as it becomes clear that drivers of mammalian evolution cannot be reduced to one limited set of factors (climatically-driven biome changes) but rather by a complex interplay between biotic and abiotic factors (e.g. biotic interactions, tectonics) (e.g. Domínguez-Rodrigo, 2014; Richerson et al., 2008).

Hereafter fossil primate taxa (such as those cited above) for which published evidences suggest frequent bipedalism in terrestrial context will be referred to as 'early hominins'.

Arid and seasonal climatic conditions were established in North Africa by the Late Miocene (~ 7-11 Myr ago), likely following the shrinkage of Tethys Sea (Zhang et al., 2014). This aridification is attested by eolian dune deposits in the Lake Chad basin (Schuster, 2006), continental records (Moussa et al., 2016; Novello et al., 2015) and marine records of terrestrial vegetation (Feakins et al., 2013). As early as 12 Ma, pollen record xerophytic plant taxa of the *Acacia-Commiphora* Somalia-Masai steppes and bushlands similar to those occurring today, but in varying proportions of grasses, Amaranthaceae, and total arboreal pollen (Bonnefille, 2010; Feakins et al., 2013; Liddy et al., 2016). Faunal assemblages typical of the savanna biome were also well established in Central and Eastern Africa by 7 Ma (Kaya et al., 2018). Vegetation modeling indicates that

Pliocene climatic conditions in Central and Eastern Africa could only sustain dry open savanna vegetation and/or xerophytic shrubland biomes (Contoux et al., 2013; Salzmann et al., 2008). Although the abundance of grass and tree pollen from 6 to 4.5 Ma in the marine record DSDP231 indicates somewhat wetter conditions in eastern Africa than during the Plio-Pleistocene and the modern time, forested habitats that hosted arboreal primates and early hominins *Sahelanthropus tchadensis*, *Orrorin tugenensis*, and *Ardipithecus ramidus* were not widespread over the northern Africa megaregion (10–30°N) (Bonnefille, 2010; Liddy et al., 2016).

In arid regions, the heterogeneity of vegetation at the landscape scale is increased by the surface water (rivers, lakes) and groundwater systems (springs), which may sustain forested areas, locally. Savanna vegetation is patchy across scales ranging from 10s of meters (local site scale) to 10s of kilometers (landscape scale) (Caylor and Shugart, 2006). This is particularly true where, despite the regional arid climate that characterizes the East African Rift Valley, rivers, shallow aquifers, and groundwater discharge areas (base level seepage and/or springs flowing from fractured bedrock), trigger the local development of azonal woody vegetation, so called spring or groundwater-fed forests and woodlands (e.g. Greenway and Vesey-Fitzgerald, 1969).

The forested habitats interpreted for *Sahelanthropus tchadensis*, *Orrorin tugenensis*, and *Ardipithecus ramidus* were sustained by local (edaphic) rather than regional (climatic) humidity. In the Toros-Menalla paleontological area (Lake Chad basin, between 7.5 and 7 Ma) the inferred paleolandscape and paleovegetation contemporaneous of *Sahelanthropus* was a mosaic of diversely wooded habitats such as dense forest patches, palm groves, and mixed/grasslands (including aquatic grasslands), with forested habitats being most likely riparian forests sustained by local surface and groundwater fed systems, rather than by regional (climatic) humidity (Contoux et al., 2013; Novello et al., 2017; Vignaud et al., 2002). According to Novello et al. (2017), “the vegetation reconstructed at the type locality of *S. tchadensis* (TM266) is similar to modern palm grove formations with an arboreal cover percentage of at least 40%”, but as noted “It is

however difficult to assess whether this particular palm grove vegetation was really the preferred habitat of *S. tchadensis* or merely a random depositional environment in this complex landscape.” At Lukeino (Kenya, ~6 Ma), faunal assemblages and isotope data on herbivorous mammals indicate that the paleolandscape contemporaneous of *O. tugenensis* included open grassy woodlands as well as patches of forested habitats likely fringing the lake margin and streams that drained into the lake (Pickford and Senut, 2001; Roche et al., 2013; Senut, 2006). At Aramis (Middle Awash Valley, ~4.4 Ma), the inferred paleoenvironment contemporaneous of *A. ramidus* was also a mosaic of habitats. It included spring-fed forest patches and woodlands that graded into more open grasslands (Suwa and Ambrose, 2014; WoldeGabriel et al., 2009).

Groundwater discharge is a common phenomenon that occurs today throughout the East African Rift (see fig.1 in Cuthbert et al., 2017). It occurred in the Aramis Member paleolandscape, 4.4 Ma ago in Ethiopia (WoldeGabriel et al., 2009), as well as in Tanzania (Ashley et al., 2016, 2010a, 2010b; Barboni, 2014) and in Kenya (Ashley et al., 2004, 2002; Johnson et al., 2009; Owen et al., 2004). Cuthbert et al. (2017) showed that in East Africa, permanent springs are most abundant in the rift, and that more than 30% of these springs were likely to remain persistent through time even if the time needed for recharging the aquifers exceeds a precession cycle of 23,000 yrs. This time may be much shorter though; for example the groundwater seeping out in the Awash Valley today in the Filwoha area of Awash National Park is only 4500-5700 years old (Bretzler et al., 2011). Cuthbert’s modeling study, which couples a hydrogeological model with an agent-based model of hominin movement also, suggests that springs in the East African Rift likely favor north – south dispersal within the rift during dry periods, and act as hydro-refugia during driest periods. Hominin survival and dispersal, therefore, may be related to the presence of springs, and more generally to the rift hydrogeological context and hydrographic network (Cuthbert et al., 2017). Although the connection between hominin or archeological sites with water may appear as an evidence for some, few studies have actually analyzed the

potential impact of springs on archeological sites and hominin behavior. Ashley et al. (2009) and Deocampo and Tactikos (2010) showed that spatial concentration of stone artifacts and butchered bones are highest at spring sites, between ~1.8 and 1.75 Ma ago, and subsequent analyses at Olduvai confirm this pattern in older paleosurfaces (e.g. Arráiz et al., 2017; Ashley et al., 2010a; Egeland, 2014). Most recently, McCool (2018) showed that groundwater discharge in the Nile River was a more reliable water source for cultural groups in the valley during the early to mid-Holocene, than surface water. To our knowledge, however, there is no account of the number of archeological and hominin sites associated with groundwater-fed systems in Africa.

Phytolith analyses carried out at the early hominin sites of Toros Menalla (Chad) and Aramis (Ethiopia) document the presence of palms (Arecaceae), which are associated with geological evidence for groundwater at Aramis, but not at Toros Menalla (Novello et al., 2017; WoldeGabriel et al., 2009). At Aramis, the pollen record narrows down the palm identification to *Hyphaene* (WoldeGabriel et al., 2009). Palms were also found associated with evidence for groundwater at several sites of the Olduvai Gorge area (Tanzania) such as in Bed I paleosurfaces sampled at FLK Zinj-PTK-AMK and FLK N immediately below Tuff IC and Tuff IF, respectively (Albert and Bamford, 2012; Arráiz et al., 2017; Barboni et al., 2010). Palm phytoliths at Olduvai were also found associated with fluvial facies at HWK W and FLK S localities within the Tuff ID/IE interval, and a leaf imprint indicates the presence of *Phoenix reclinata* (Albert et al., 2018). The present-day distributions of *Hyphaene* and *Phoenix* species are different. Their occurrence in the fossil record may therefore indicate particular paleoenvironments. Groundwater discharge areas leave evidence in the geological record (e.g. tufa and carbonate layers with freshwater isotopic signal) (Ashley et al., 2014b, 2010b, 2010a). They potentially leave evidence in the paleovegetation micro-botanical record as well, but this has barely been analyzed to date.

The aim of this paper is to explore importance of groundwater, and of groundwater-associated woodlands and forests

to the record of hominins. It provides new data (field observations, pollen and phytolith data) on modern groundwater systems and their associated microhabitats from two regions in Eastern Africa where hominin and archeological sites are numerous. It provides a summary of the groundwater-fed paleoenvironment inferred for the *Ardipithecus*-bearing Lower Aramis member of the Central Awash Complex of the Middle Awash area (Ethiopia), the most intensively studied groundwater-associated hominin site. A survey of other hominin and archeological sites found in association with spring sites in Africa and the Middle East is also presented. The importance of groundwater-fed systems for species and ecological processes is discussed, as well as what it may imply for early hominins such as *Ardipithecus ramidus* at Aramis, 4.4 Ma ago.

## 2. Modern springs in East Africa: some case studies from Ethiopia and Tanzania

Fieldwork was carried out in Ethiopia and Tanzania, at modern spring sites, to document vegetation patterns, groundwater characteristics, geologic context, and to collect surface soil samples for phytolith and pollen analyses.

### 2.1. Field observations on geomorphology and vegetation

Springs and groundwater-fed wetlands in the East African Rift are generally found at the base of slopes and associated with faults (Cuthbert and Ashley, 2014; Olago et al., 2009). In the Awash River valley (Ethiopian Afar rift, Ethiopia), the three spring sites we visited occur in grabens, at the base of rift shoulders (between 750 and 550 m asl) (Fig. 1). Spring recharge occurs in the Western highlands and on nearby volcanoes at high elevation (> 2900 m asl). The slopes of the aquifers are important because the distance between the rift western flank where recharge occurs and grabens where springheads are located is short (50–90 km) (Fig. 1C). This part of the rift valley is characterized by tectonism and four recently active volcanoes (Chorowicz, 2005). At the three Ethiopian sites spring water was hot (>40–50°C) and pH about 8–9. In the Lake Eyasi – Manyara region (Gregory Rift, north Tanzania), the spring sites visited occur in grabens and at the base of slopes of the Ngorongoro Crater Highlands (Fig. 2).

Spring recharge occurs on the Ngorongoro Crater Highlands (>3000 m asl) located nearby (15–25 km from spring lines). At Lake Manyara and Lake Eyasi (Kisima Ngieda area) where large palm woodlands and wetlands occur (Fig. 2B), spring water is fresh (16–22 °C) and pH between 5 and 8. Some springs also occur on the western edge of Lake Manyara and northern and eastern edge of Lake Eyasi.

Depending on the geological context and/or on the spring discharge flux, we observed different spring habitats: herbaceous wetlands where groundwater table reached the surface (Fig. 1D, 3D, 4E), palm groves where groundwater was saline/alkaline and ca. 100 cm below the surface (Fig. 3, 4B, 4F-G), and evergreen forests where groundwater was fresh and most abundant (Fig. 4A, 4C). Vegetation at the spring sites we visited is azonal. It is strikingly different from vegetation in the surrounding landscape (Fig. 1D, 2C). The modern groundwater forest that occupies the northern end of saline Lake Manyara (Tanzania) is an excellent example of azonal vegetation (Fig. 2C, 4A). Plant species growing under the spring influence, make a dense forested patch of tall, evergreen trees, providing shade, cooler microclimate and micro-environment that are strikingly different than the nearby bushland (Fig. 4D), where deciduous spiny trees and shrubs, generally of low height except baobabs, offer little shade and protection against the heat of the day (Greenway and Vesey-Fitzgerald, 1969; Loth and Prins, 1986).

At groundwater-fed wetlands we observed abundant aquatic herbs *Typha*, *Phragmites*, ferns and Cyperaceae species, and, some hydrophyte treelets such as *Sesbania sesban* (in Tanzania near Lake Eyasi) or *Tamarix* (in the Awash Valley). At spring sites where groundwater is very close to the surface (ca. 100 cm) we observed palm woodlands. Two genera, *Hyphaene* and *Phoenix* were commonly found at spring sites. In the Awash Valley, *Hyphaene thebaica* (doum palm) was found at Doho, and in the Filwoha area of the Awash National Park where it forms large groves over an area of several square kilometers (Fig. 3A, 3C). At Metaka, where the spring water merges with the Awash River, only a single *Hyphaene* tree was present. No palm trees were observed at Bilen, likely because of logging. In Tanzania, another species, *Hyphaene petersiana* was observed on the northeastern side of Lake Eyasi where it

forms a ~15 km-long woodland in association with the yellow bark *Acacia xanthophloea* (Fig. 2B). *Hyphaene petersiana* and *Phoenix reclinata* were both observed on the northwestern edge of Lake Manyara; *Phoenix* occurring under the tree canopy of the spring forest, while *Hyphaene* occur on open, dry grounds. In *Hyphaene* palm groves, we note that the abundance of grasses (Poaceae) in the understory is striking; see example from the Filwoha National Park in Ethiopia (Fig. 3A). Other characteristic trees occur at spring sites e.g. *Acacia xanthophloea*, *Rauvolfia caffra*, and *Tamarindus indica* in Tanzania, *Tamarix*, and *Ficus* spp. in Ethiopia, but *Typha*, Cyperaceae, Poaceae and palms of *Hyphaene* or *Phoenix* species are common to all sites. Palm groves and islands of *Phoenix reclinata* have also been observed along the Banagi and Bonar groundwater-fed rivers in the western Serengeti (Albert et al., 2015, and Ashley personal observations).

*Hyphaene* palm groves seem to develop where groundwater is saline and/or alkaline, whatever the groundwater temperature (cool or hot). We observed that *Hyphaene petersiana* occurs on dry grounds, but where water table is high (about 100 cm below the surface), at the ecotone between saline/alkaline Lake Eyasi and the Kisima Ngieda freshwater spring and wetland, in northern Tanzania. In the Awash Valley, *Hyphaene thebaica* may occur directly at the springheads, growing within the fractures of the basalt (Fig. 3D). Hence, the presence of *Hyphaene* at spring sites is linked to high water table and relatively high water salinity or alkalinity (Stauffer et al., 2014). Spring water temperature, which we measured to be up to 50°C does not seem to be a limiting factor for *Hyphaene*. *Phoenix reclinata*, on the contrary to *Hyphaene* species, favors less saline/alkaline but wetter (damp) soils and may occur in both open and shady habitats.

## 2.2. Pollen analyses

Pollen data are available for 20 surface soil samples, which we collected at several spring sites in the Awash Valley (5 samples, labeled BB17-#) and in the Eyasi – Manyara region (15 samples, labeled DB11-#). A surface soil sample consists of 20–30 sub-samples collected at random over an area of about 500 m<sup>2</sup> and combined. Soil samples (without litter) were collected 0 to 1 cm below surface following

classical sampling method (Bonnefille et al., 1999). Extractions and counting procedures also follow Bonnefille et al. (1999).

*Modern pollen signal of groundwater forests and woodlands* (Fig. 5, SOM 1). Herbaceous taxa such as *Typha*, Cyperaceae, and Poaceae, as well as the arboreal pollen of *Acacia*, occur in all 20 samples in various proportions: Cyperaceae being the most abundant taxon followed by Poaceae, *Typha*, and *Acacia* with average relative abundances of 33%, 12%, 11%, and 3% respectively. Herbaceous taxa such as Asteraceae, *Cyathula*-type *orthocantha*, *Achyranthes*-type *aspera* and other Amaranthaceae also occur in all samples albeit low relative abundances (<1-5%). In agreement with field observations, *Hyphaene*-type is most abundant in the palm woodlands from the Awash Valley and Lake Eyasi, while *Phoenix reclinata*-type is most abundant in samples from the Manyara spring-forest. Pollen taxa diversity is highest in the Manyara spring forest. Arboreal pollen taxa, palms excluded, however, never represent more than 35% of the total pollen assemblages, even in the Manyara forest samples. We note that *Celtis*, *Syzygium*-type *guineense* and *Ficus*-type (found as fossil seeds and wood in Aramis Member) only occur in samples from the most developed spring forest at Manyara; *Ficus*-type, however, also occurs in two samples from Eyasi palm woodlands (SOM 1).

### 2.3. Phytolith analyses

Samples analyzed for their pollen content were also analyzed for their phytolith content, except BB17-17. Extractions and counting procedure follow WoldeGabriel et al. (2009). The phytolith data obtained for these 14 samples were included in the African modern soil phytolith dataset, which initially was made up of 149 samples (Barboni et al., 2007; WoldeGabriel et al., 2009). In this previous dataset, not a single sample came from a groundwater-fed zone. We have augmented this dataset with 116 new data points including data from various African vegetation types where palms are well-represented such as in some West African gallery forests, desert oases, riparian palm islands from groundwater-fed rivers, and the spring-associated woodlands and forests (Albert et al., 2015; Arraiz, 2017; Novello, 2012; Novello et al., 2017). In addition to these, we included samples from zonal

vegetation types in Guinea, Chad, and Tanzania, which are more distinctly influenced by climatic than by edaphic factors. Fossil phytolith data such as Aramis can now be compared against a modern African dataset of 265 samples (SOM 2).

We have chosen to constrain our interpretations to phytolith morphotypes with clear taxonomic attribution, namely silica shorts cells for grasses (rondel, bilobate, polylobate, cross, saddle and crenate morphotypes) (e.g. Barboni and Bremond, 2009), globular decorated morphotypes for woody dicotyledons (representing trees and shrubs essentially) (Collura and Neumann, 2017), and globular echinate morphotypes for palms (e.g. Bamford et al., 2006). However, the rondel morphotype was excluded from the correspondence analysis (CA) as it prevented sample discrimination according to vegetation types, as also observed by Neumann et al. (2017). CA was run on raw phytolith counts, using modern surface soil samples as active variables, and Aramis samples (SA-#) as supplementary variables. We used R software using FactoMineR (for the analysis) and factoextra (for data visualization) (Kassambara and Mundt, 2017; Lê et al., 2008; R Core Team, 2018) (Fig. 6).

*Modern phytolith signal of groundwater forests and woodlands* (Fig. 6, SOM 2). Correspondence analysis of the African surface soil phytolith assemblages shows that Axis 1, which explains 29.2% of the total inertia, distinguishes forests and woodlands from grasslands. Axis 2 (20% of total inertia) separates low elevation from high elevation grasslands on the one hand, and, on the other hand, brings some discrimination among the wooded environments. Dispersion of the phytolith morphotypes about Axes 1 and 2 indicates that desert oases, *Raphia* swamp, groundwater forests and palm woodlands are characterized by the globular echinate phytoliths, while semi-deciduous forests and evergreen forests are characterized by phytolith assemblages with abundant globular granulate and smooth morphotypes. The cohort of phytoliths typical for lowland grasslands, the bilobate, cross, and saddle grass silica short cell morphotypes, are also typical for C4 grasses, while the crenate morphotypes is characteristic for high elevation grasslands where C3 Pooideae grasses dominate (Barboni and Bremond, 2009).

### 3. Fossil springs

#### 3.1. Aramis Member, Awash Valley, Ethiopia

The *Ardipithecus*-bearing Lower Aramis member of the Central Awash Complex of the Middle Awash area (Ethiopia), dated 4.4 Ma, was thoroughly documented and provides the most complete set of geological, faunal, isotopic, and botanical data associated with the early hominin *Ardipithecus ramidus* (Louchart et al., 2009; Suwa et al., 2009; White et al., 2009b, 2009a; WoldeGabriel et al., 2009). From our point of view, it also provides a thoroughly documented multi-proxy record of a paleo-ecosystem associated with groundwater.

Geological evidence for groundwater in the Aramis member are described as follows by WoldeGabriel et al. (2009), which make no doubt that groundwater strongly influenced botanical and faunal assemblages:

"Massive (<1.5 m thick), predominantly micritic carbonate horizons and nodules representing groundwater and pedogenic deposits pinch out laterally within clayey silts. These are also locally fossiliferous. Carbonate deposits in some localities contain characteristic features of tufas, such as fossil gastropods and other invertebrates, abundant and uncrushed calcite-replaced vegetation, vertebrate remains, and eggshells (guinea-fowl size). These suggest that the carbonate horizons generally formed at or near the landscape surface. Evidence of spring activity includes several 1-m-wide banded travertine deposits associated with faults. A porous microcrystalline carbonate with dense concentrations of calcite isomorphs of plant parts forms a broad, low dome just north of ARA-VP-6. However, in almost all sections excavated for isotopic, phytolith, and pollen analysis, the carbonates lack diagnostic features of tufas. Their micritic textures and the presence of terrestrial soil invertebrate faunal activity (such as dung beetle brood burrows) suggest that the carbonate horizons are derived from groundwater carbonate that generally formed at or near the landscape surface in seasonally saturated soils near springs".

Over the 7 km-long west to east transect sampling Aramis paleolandscape, carbonate horizons are found at ARA-VP-17 and ARA-VP-6 localities capped by the Daam Atuu Basaltic Tuff, and about 70-80 cm below the DABT at KUS-VP-2 and SAG-VP-7 localities (WoldeGabriel et al., 2009, figure 1).

Calcite-replaced wood and endocarps were found preserved in the carbonate horizons of Aramis Member (mainly at ARA-VP-6 locality), and absent in the non-cemented sediments (WoldeGabriel et al., 2009). In agreement with geological and faunal evidence for negligible fluvial transport, the wood and seed specimens attributed to *Ficus*, *Syzygium cf. guineense*, and *Celtis* (Jolly-Saad and Bonnefille, 2012), therefore, could have been fossilized *in situ* or not far from the place of collection, i.e. not far from the spring resurgence. The Aramis pollen record, despite its paucity (only 16 grains found in just 4 samples over 40 tested), attests for the presence of *Hyphaene* (n = 2 at ARA-VP-6 and n = 2 at ARA-VP-1 TS) (WoldeGabriel et al., 2009). *Hyphaene* pollen type is produced by both *Borassus* and *Hyphaene*; however, *Borassus* palm trees do not occur at spring sites but in dryland savannas, on the contrary to *Hyphaene* species which grow at alkaline spring sites, preferentially on rocky grounds, and where water table is high (Edwards et al., 1997; Orwa et al., 2009). The presence of palms at Aramis is also attested by the phytolith record in which globular echinate phytoliths may represent up to 40% (WoldeGabriel et al., 2009). Pollen, relatively prone to transport, indicate that trees of *Myrica* and of *Hyphaene* palms were also part of the vegetation locally or in the landscape along with grasses and sedges (WoldeGabriel et al., 2009). *Myrica* species today occur in the Afromontane forests, while *Hyphaene* occur in the lowlands (up to 1400 m) (Orwa et al., 2009).

Aramis phytoliths assemblages indicate the presence of C<sub>4</sub> grasses, palms, woody plants, and sedges. Phytolith-inferred woody cover is heterogeneous among samples as it ranges from <40 % to ~65 % (WoldeGabriel et al., 2009). By comparison with our extended modern surface soil phytolith dataset, Aramis fossil phytolith assemblages (SOM 3) best compare with present-day lowland C<sub>4</sub>-grasslands, groundwater forests and palm woodlands. Few samples (SA15, SA19, and SA39) show similarities with semi-deciduous forests and the

group of samples that includes Tree/Shrub savannas, ecotones, and the Awash riparian forest (Fig. 6).

Aramis paleosol carbonates, which were sampled over the Aramis 7 km-long west to east paleolandscape transect, exhibit  $\delta^{13}\text{C}$  values ranging from -9.3‰ to -0.4‰, and  $\delta^{18}\text{O}$  values ranging from -9.8‰ to -0.9‰ (WoldeGabriel et al., 2009) (reported in Fig.7). They indicate expanses of wooded grassland (tree or bush-savanna) where densely wooded habitats were likely of limited geographic extent (Cerling et al., 2010; WoldeGabriel et al., 2009). Wooded habitat, however, was inferred from multiple lines of evidence, such as the abundance of *Tragelaphus* (browsing antelope, kudu) and cercopithecid monkeys (*Pliopapio alemui* and *Kuseracolobus aramisi*) which combined with other data show that “the large mammal biomass at Aramis was dominated by browsers and frugivores” (White et al., 2009a). Preferred wooded habitat was clearly inferred for the primates *Ardipithecus ramidus*, *Kuseracolobus aramisi* (colobine monkey) and *Pliopapio alemui* (a small baboon-like monkey), and *Tragelaphus* using carbon isotopes in enamel (Fig.7), tooth micro- and mesowear, craniofacial structure and masticatory apparatus, and tooth anatomy and proportions for the hominin, and microwear, mesowear, isotopes, and postcranial ecomorphology for cercopithecids and kudus (White et al., 2015a, 2009a). The micro-mammal assemblage indicates a variety of biotopes ranging from locally present forests and/or well-developed mesic woodlands, to palm thickets/woodlands, savanna woodlands, dry scrub or even arid steppe. The assemblage also includes two taxa which present-day counterparts occur in mesic montane forests and uplands (Louchart et al., 2009). Regarding the bird assemblage, we note that among the Psittacidae, Apodidae and Passeriformes, which represent >43% of the identified avian specimens at Aramis (Louchart et al., 2009), several have modern species in Ethiopian lowlands that rely on the fruit bearing spring palm species *Hyphaene thebaica* (Ash and Atkins, 2010). At last, the terrestrial gastropods assemblage with *Maizania* from the *M. hildebrandti* group, *Limicolaria* sp. and *Chlamydarion* cf. *hians* bears resemblance with that of modern Kibwezi groundwater forest in southern Kenya (WoldeGabriel et al., 2009).

### 3.2. Hominin and archeological sites co-occurring with springs

Springs and groundwater-supported ecosystems may leave sparse evidence in the geologic record. Yet, evidence left is diverse and corresponds to *in situ* deposits, so important for paleoenvironment interpretation. They include terrestrial and aquatic plant remains such as pollen, phytoliths, seeds and other carbonated and silicified macro-remains as described in length here for Aramis Member, as well as algae (Ashley et al., 2016), and organic molecules (Magill et al., 2016). Evidence also include *in situ* terrestrial and aquatic animals (mollusks, crustaceans) (e.g. Pickford, 1995), and *in situ* minerals: chiefly carbonate, but also opal (silica), evaporates, sulphates, clay minerals (Cantonati et al., 2016; Pigati et al., 2014).

Copious literature documents hominin and/or archeological sites associated with evidence for groundwater, with no less than 50 different localities within 22 geographically different paleontological and archeological areas found throughout the East African Rift and beyond in the Arabian Peninsula, Jordanian, Iran, and Turkey (Fig. 8, Table 1). Among these, 12 are dated to > 1 Ma. In some cases, springs and hominin remains and/or artifacts may be found to co-occur in several stratigraphic levels within one paleontological area. At Olduvai Gorge for example, springs and hominin remains and/or artifacts were found to co-occur in no less than five different paleosurfaces between 1.85 and 1.34 Ma (Table 1).

Interestingly for the Pliocene, both fossiliferous areas in the Afar region with *Ardipithecus ramidus*, Gona and Aramis, have spring deposits (Semaw et al., 2005; WoldeGabriel et al., 2009). Despite different depositional environments at Gona and Aramis, *Ardipithecus* remains were found closely associated with tufa or groundwater carbonates. Spring deposits also occur at several sites with *Australopithecus* species in Ethiopia (Woranso-Mille) (Haile-Selassie et al., 2007), South Africa (Taung) (Hopley et al., 2013; McKee and Kuykendall, 2016) and plausibly in Tanzania as well. At Laetoli (Tanzania), geological evidence for groundwater was likely eroded although “pond deposits” were recognized (Ditchfield and Harrison, 2011, p. 74). We also suspect the presence of springs at Laetoli

because pollen taxa such as *Hyphaene* (spring palm tree) and the freshwater-loving *Typha* (cattail) occur in several samples (Barboni, 2014; Bonnefille and Riollet, 1987).

For the early Pleistocene, spring sites with hominin remains are found in Turkey (Kappelman et al., 2008; Lebatard et al., 2014; Viallet et al., 2012), the Syrian desert (Jagher et al., 2015), and in Tanzania with several sites throughout the 1.89 – 1.30 Ma interval at Olduvai Gorge (Ashley et al., 2016, 2014a, 2010a, 2010b, 2010c, 2009; Barboni et al., 2010; Deocampo et al., 2002; Garrett, 2017; McHenry et al., 2007), and possibly at Peninj as *Typha* pollen is abundant (Dominguez-Rodrigo et al., 2001). During the Middle and Upper Pleistocene, there is recurrent evidence for human presence at spring sites in the Saharan desert belt (Churcher et al., 1999; Dachy et al., 2018; Foulds et al., 2017; Hill, 2001; Inglis et al., 2017; Kleindienst et al., 2008; McCool, 2018; Nicoll et al., 1999; Smith et al., 2007, 2004; Wendorf et al., 1993), in northeast Ethiopia (Benito-Calvo et al., 2014; Gossa et al., 2012; Williams et al., 1977), in Kenya (Beverly et al., 2015; Johnson et al., 2009; Johnson and McBrearty, 2012; Tryon et al., 2014, 2012; Van Plantinga, 2011), and South Africa (Butzer, 1973; Porat et al., 2010). In South Africa, the massive tufa fan deposits spanning the length of the Ghaap Plateau escarpment (the Buxton Limeworks at Taung, and the Groot Kloof and Gorrokop at Ulco) have resulted from the discharge of groundwater-fed Thabaseek River since the Pliocene (review in Doran et al., 2015). The Buxton Limeworks are aggrading surface freshwater carbonate deposits at the edge of the Kalahari Desert that have preserved the Taung Child skull attributed to *Australopithecus africanus* (Dart, 1925), as well as traces of at least 17 Pleistocene and Holocene sites (Hopley et al., 2013; McKee, 1994; McKee and Kuykendall, 2016).

#### 4. Discussion

##### 4.1. Modern analog groundwater-associated habitats

The pollen signal of the groundwater-supported habitats we sampled is characterized by the abundance of herbaceous taxa such as *Typha*, Cyperaceae, and Poaceae, while the abundance of arboreal taxa is low (<~35%),

except for *Acacia*, and palms *Hyphaene* and *Phoenix*. The Manyara forest for example, despite its closed, mostly evergreen canopy and important tree taxa diversity is characterized by low percentages of arboreal pollen, because Cyperaceae (sedges), which occur in enclosed swamp herbages (Fig. 5) contribute to disproportionately high pollen percentages (Fig. 6). Among the groundwater-associated habitats, therefore, forests may not show up in the fossil pollen record, which is likely to exacerbate the signal of wetlands, as it is dominated by herbaceous aquatic taxa. Woodlands and forests may also be outshined in the pollen record by the surrounding open and more largely widespread xerophytic vegetation, as some xerophytic taxa, notably Amaranthaceae occur in all surface samples, even in most forested sites. We note that although *Ficus*, *Celtis*, and *Syzygium* arboreal taxa were recorded in the pollen assemblages from the most developed forest of Manyara, these taxa are not unique to groundwater-fed forests as they also occur in riparian forests (Carr, 1998).

In the phytolith record, some, but not all groundwater-associated habitats carry a distinct phytolith signal. Wetlands, on the one hand, do not differentiate despite the fact that Cyperaceae produce typical (unique) morphotypes; the silicified papillae phytoliths of sedges were too poorly represented in the soil phytolith assemblages to allow discrimination (Fig. 6). Wetlands exhibit a phytolith signal indistinguishable from that of grasslands most likely also because our dataset was restricted to main phytolith categories. Novello et al. (2012) showed that distinguishing trapeziform grass silica short cell phytoliths within the bilobate, cross, and saddle categories could improve the identification of wetland grasses. Unfortunately, such distinction was not considered by all the authors who contributed to the African phytolith dataset presented here. Some groundwater-associated woodlands and forests, on the other hand, strongly discriminate because of the presence and relative abundance in the surface samples of globular echinate phytoliths, typical for palms (Arecaceae). It is the presence of palms in the vegetation that is well captured by the phytoliths and which, therefore, allows identifying some groundwater-associated woodlands and forests, such as the Manyara spring forest, the Eyasi

*Acacia* and palm woodland, a palm grove on a spring-fed river bank in the Serengeti, the Saharan oases, and some palm woodlands from the Awash River valley. We note, however, that despite the presence of palms in the vegetation, the relative abundance of globular echinate phytoliths was too low in several samples to allow discriminating some groundwater-associated palm woodlands: e.g. samples BB17-05 from a *Hyphaene thebaica* palm woodland in the Awash Valley, and MNY12-40 from Manyara forest (Fig.6). The systematic over-representation of palms in phytolith assemblages is, thus, not proven here. Other authors came to the same conclusion: although Arecaceae are large phytolith producers (Hodson et al., 2005), palms are not systematically over-represented in surface samples (Albert et al., 2015; Bremond et al., 2005; Novello, 2012; Novello et al., 2017). Yet, to our knowledge, a formal calibration between phytolith abundance and palm abundance in the vegetation has never been carried out.

Taphonomic issues affect phytolith assemblages, and dissolution affects phytolith morphotypes differently despite the fact that they all are composed of the same mineral ( $\text{SiO}_2$ ,  $n\text{H}_2\text{O}$ ) and have similar range of solubility (Fraysse et al., 2009, 2006). Phytolith dissolution starts at  $\text{pH} > 8$ , and preferentially affects morphotypes with a surface to bulk ratio  $> 1$ , such as e.g. the silicified papillae (hat-shaped phytoliths) of Cyperaceae. Morphotypes with a surface to bulk ratio  $< 1$ , such as the grass silica short cells and the globular echinate phytoliths of palms happen to be particularly stable (Cabanes and Shahack-Gross, 2015). In the 13 fossil phytolith assemblages from Aramis Member where hat-shaped phytoliths occur (WoldeGabriel et al., 2009), it is therefore likely that preservation was exceptional, and that the relative abundance of all morphotypes was preserved. We agree that a standardized phytolith solubility test as proposed by Cabanes and Shahack-Gross (2015) would allow evaluating the phytolith state of preservation for the whole Aramis dataset.

#### 4.2. Identifying ancient groundwater-associated habitats in the geologic, isotopic, and faunal record

The groundwater-discharge zones are typically localized features that can be as small

as 10s of  $\text{m}^2$  or a few kilometers<sup>2</sup> (Magill et al., 2016; Pigati et al., 2014). Groundwater-discharge zones, thus, could be completely missed when sampling paleoenvironments. Among the micro-habitats supported by groundwater, the least extensive are the forests, as they are the most demanding in terms of water amount, quality, and availability through time. Hence, although springs may be permanent features on the landscape lasting for hundreds of years, groundwater-supported wetlands and woodlands may leave only sparse evidence in the geologic record. At Aramis (4.4 Ma), for example, the woodland setting with forest patches that made the habitat of kudus and leaf- and fruit-eating primates including *Ardipithecus*, did leave a forest signal in the tooth enamel isotopic record of browsing taxa (White et al., 2009a), but not so in the pedogenic carbonates, as most  $\delta^{13}\text{C}$  values cluster between -5 ‰ and -2 ‰ (WoldeGabriel et al., 2009). The discrepancy between the enamel and pedogenic carbonate isotopic datasets (Fig. 7) may be related to the fact that isotopic records of tooth enamel captures the herbivores' dietary preference over the herbivores' lifetime (1–10 years), while that of pedogenic carbonate averages environmental information over 100–1000 years (Du et al., 2019). It could also be related to the fact that forest patches at Aramis were too sparse compared to open (grass-prone) habitats in the landscape that was sampled, to significantly contribute to the isotopic record in carbonate nodules (Cerling et al., 2011). An alternative possibility, more convincing to our opinion, is that the carbonate nodules sampled in the Aramis member were diagenetically altered by groundwater, and that their carbon isotopic composition was overprinted by the groundwater own isotopic composition (Budd et al., 2002; Gallagher and Sheldon, 2016). In the case where the presence of groundwater is attested,  $\delta^{13}\text{C}$  should be measured on organic biomarkers (leaf waxes) rather than on carbonate nodules, as they are too prone to having their  $\delta^{13}\text{C}$  values reset by the groundwater carbon isotopic composition and the extent of water:rock interaction (Budd et al., 2002).

Groundwater-fed environments have tended to be overlooked and underappreciated in the geological record. Similarly, in the modern world, groundwater-fed habitats have

hardly been sampled to evaluate their proxy signature. The modern phytolith dataset initially used to interpret Aramis fossil data did not include any samples from groundwater micro-habitats (Barboni et al., 2007; WoldeGabriel et al., 2009), nor did the isotopic dataset assembled to provide a quantitative modern-based estimate of paleo tree cover at Aramis (Cerling et al., 2014, 2011, 2010). In this modern isotope dataset only one sample evaluates the isotopic signal of a groundwater-forest in Kenya (Mzima Springs, Tsavo West National Park), but not a single sample comes from palm-rich groundwater-associated woodlands, the most likely analog according to the whole Aramis dataset. The paleoenvironment of *Ardipithecus ramidus* at Aramis was interpreted as a distal floodplain with spring-fed forest patches within predominantly grassy woodland to wooded grassland habitats (Suwa and Ambrose, 2014; White et al., 2009b; WoldeGabriel et al., 2009). Yet in a discussion of the White et al. interpretation, Cerling et al. (2014, 2010) interpreted a tree- or bush-savanna and did not recognize the possibility of localized water sources unrelated to rivers and lakes. Cerling et al. (2014, 2010)'s interpretations were of a generalized landscape that was dry everywhere except perhaps along a riparian corridor. Yet, Aramis is not in a riparian setting (White et al., 2015b; WoldeGabriel et al., 2009). The importance of groundwater in providing relatively persistent additional moisture, therefore, still needs to be understood and acknowledged.

The Aramis groundwater-fed paleoecosystem included the early hominin *Ardipithecus ramidus*, as well as a diverse fauna of macrovertebrates characterized by the abundance of tragelaphine bovids and cercopithecid primates (White et al., 2009a). Some bovid tribes are useful environment indicators: the grazing Alcelaphini, Antilopini and Hippotragini (AAH) indicate open grass-dominated habitats, the mixed feeders and browsing Tragelaphini and Aepycerotini (TA) indicate dry woodlands, and Reduncini and Bovini (RB) are usually associated with closed riparian habitats (Shipman and Harris, 1988; Vrba, 1980). At Aramis, tragelaphines represents 85%, aepycerotines 4%, reduncines <1% and all other tribes less than 10% of the

Bovidae. Aramis bovid assemblage does not compare to Olduvai Bed I and Bed II (Alcelaphini and Antilopini-rich), despite the recognized presence of groundwater-fed wooded and palm-rich micro-habitats (Albert et al., 2018; Arráiz et al., 2017; Ashley et al., 2010b; Barboni et al., 2010), nor to Shungura (mixed Reduncini and Tragelaphini) where fluvial and deltaic environments likely prevailed (Shipman and Harris, 1988). Aramis bovid assemblage is very different than many other Plio-Pleistocene sites by its marked abundance of tragelaphines (White et al., 2009a). The dominance of Tragelaphini at Aramis best compares with modern Kruger, Mkuzi and Timbavati national parks (N.P.) bovid data when comparison is restricted to AA, BR and TA tribes (Aramis data plotted on Shipman and Harris, 1988's ternary diagram, not shown). By considering all bovid tribes but Reduncini, however, Aramis data best compare with Manyara N.P. and Hwange N.P. (Domínguez-Rodrigo and Musiba, 2010). Manyara N.P. is definitely a groundwater-fed ecosystem, as described in length here. Hwange N.P. includes many spring areas (e.g. Sinamatella), and could therefore represent another potential faunal analog for Aramis.

In Aramis groundwater-fed ecosystem primates in the Cercopithecidae include a colobine monkey (*Kusleralocolobus aramisi*) and a small baboon-like monkey (*Pliopapio alemui*) (White et al., 2009a). Today, the Filwoha spring area in the Awash Valley doesn't host colobines, but Hamadryas baboons (*Papio hamadryas hamadryas*), which large group sizes may be explained by "the abundance of one food resource in particular, doum palm nuts" (i.e. *Hyphaene thebaica*) (Swedell, 2002). One colobine, *Colobus guereza* is found today in several Ethiopian regions largely at mid- and high-elevation (>1700 m up to 3300 m asl) and at low elevation in the Omo Valley and, presumably, the Awash Valley (Dunbar, 1975). This leaf-eating monkey, however, chiefly occupies riparian and gallery forests where its favorite food are leaves of *Celtis africana* and *Ficus* spp. (Dunbar, 1987; Dunbar and Dunbar, 1974), among many other species (Hussein et al., 2017). Understanding how critical the role of groundwater-fed habitats was to the Aramis faunal community, however, would require further investigations.

#### 4.3. Why are springs ecological keystones?

Despite the fact that groundwater-fed areas leave well-recognized evidence in the geological record, their importance in defining the type of habitat available to hominins and other elements of the fauna is rarely recognized (but see Barboni, 2014; Beverly et al., 2015; Cuthbert et al., 2017; Deocampo and Tactikos, 2010; Djamali et al., 2018; Reynolds et al., 2011). Yet, we have shown here that the occurrence of springs is not anecdotal in the African hominin and human record (Table 1). Aramis site is just one among > 50 examples in Africa and the Middle East where springs or groundwater-fed areas co-occur with hominin and/or archeological remains. Here we discuss the importance of springs, and that springs are ecological keystones for species, and crucial environmental features that cannot be ignored in paleontology, and in paleoenvironmental studies aiming at reconstructing hominin paleo-habitats.

Springs in arid environments play key roles for species. In the discipline of conservation biology, small natural features like springs and riparian areas are considered ecological keystones because these sites have a disproportionate ecological importance to their size (Hunter Jr et al., 2017). Springs and groundwater-fed habitats have been recognized biodiversity hotspots in Europe (Cantonati et al., 2012), North America (Stevens and Meretsky, 2008), New Zealand (Collier and Smith, 2006), as well in tropical deserts in Africa (Suhling et al., 2006), Australia (Davis et al., 2017; Fensham et al., 2011; Murphy et al., 2015), and Central America (Boggs et al., 2014). In xeric regions, springs provide permanent source of water for vertebrates, and are used as mesic refugia by birds (e.g. Szaro and Jakle, 1985), elephants (e.g. Viljoen et al., 1990), as well as savanna chimpanzees (Kempf, 2009; Pruetz and Bertolani, 2009).

##### 4.3.1. Springs provide a stable water resource

Groundwater-discharge zones may supply the critical need for water for plants and animals in arid regions, as springs persist throughout the year and throughout wet-dry climatic cycles. The presence and temporal persistence of springs depends on multiple factors. Cuthbert et al. (2017) showed that groundwater response

time is the primary factor controlling the presence and the persistence of active springs in the East African Rift, rather than rates of groundwater recharge. Groundwater response time depends on subsurface hydraulic properties of the aquifer and the topography (i.e. length scale and topographic gradient of the catchment area), two factors that are stable over  $10^3$  to  $10^6$  years-long periods. On the contrary, rates of groundwater recharge depend on climate, a factor that varies greatly over a range of timescales (from  $<10^1$  to  $>10^5$  years-long periods). Hence, topography and geology act as buffers; they prevent springs from responding directly to climate changes (Cuthbert et al., 2017). Springs may therefore guarantee a rather stable habitat. They provide ecological continuity through time. At Esere, near Laetoli paleo-anthropological area (northern Tanzania, Fig. 2B), an active spring occurs next to a ca 2 m-thick carbonated tufa mount attesting for the presence of a persistent fresh water spring in this area (dating of the carbonate in progress, Ashley, *unpublished*) (Fig. 9). Springs were likely present throughout the Late Pliocene – Early Pleistocene in this region of north Tanzania, where no less than three paleontological sites are known (Barboni, 2014).

##### 4.3.2. Springs are controlled by geology and topography, and are independent of climate

Springs are more likely to play a crucial ecological role in regions with low annual rainfall, seasonal rainfall and periodic droughts because groundwater is protected from evaporation. At first glance, this is counter-intuitive as a reasonable assumption has been that more rainfall (recharge) would lead to more discharge (spring and groundwater seeps) and that climate variability is the dominant control on water availability. But, the persistence of springs is highly dependent on topographic and geologic factors, such as distance from recharge area, the topographic gradient and the transmissivity of the intermediary rocks and soils (Cuthbert et al., 2017). These physical controls buffer the impact of climate variability. A majority of the paleo-spring records at Olduvai Gorge are located in topographic low areas (playa lake basin), at the base of the slope and formed during insolation minima during periods of low lake levels (Ashley et al., 2014b;

Cuthbert et al., 2017; Cuthbert and Ashley, 2014).

#### 4.3.3. Springs generate distinct microclimates, distinct vegetation and distinct micro-habitats

The sustained water supply increases structural complexity because soil moisture triggers higher plant productivity. In East Africa, rainfall is mostly low (250–500 mm/yr) and highly seasonal, whereas evapotranspiration is high (~2500 mm/yr). As a consequence, shallow lakes are saline and net primary productivity is low compared to e.g. central West Africa (Brown et al., 2010). Vegetation growth is limited in many areas, and trees are generally scattered in the landscape. It is a region where climate deterministically supports low (<55%) tree cover (Staver et al., 2011). Tree cover >55% only occurs where rainfall is >750 mm/yr and dry season <7 months (Good and Caylor, 2011; Staver et al., 2011). Springs are eye-catching in the landscape, as they appear like vegetated island oases in an otherwise grass-dominated or bare landscape. Groundwater-fed areas are more vegetated than the surroundings; a contrast that is well marked on satellite images (Figs. 2, 3) (Reynolds et al., 2016). The size of a groundwater discharge area varies according to groundwater discharge rate and the geological setting, which may favor the development of a groundwater-fed river or more expansive wetland. In Africa, the greatest number of habitats are seen near water bodies, and decline with distance (O'Regan et al., 2016).

At the local scale, springs increase soil moisture such as high-water demanding plants may grow even under arid climate. High soil moisture also favors the growth of tall trees, and evergreen (rather than deciduous) plants. Springs therefore favor the growth of azonal vegetation units, which are not in equilibrium with regional climate (e.g. Greenway and Vesey-Fitzgerald, 1969). Groundwater-fed areas support denser vegetation and taller trees than surrounding areas; the microclimate they provide contributes to buffer extreme temperatures and maintain moisture. Pruetz and Bertolani (2009) observed that spring-associated gallery forests, which provide the only permanent source of water and shade during the dry season can be considered as

“pseudo-home base for Fongoli chimpanzees at this time, as they move outwards from these areas in a radiating fashion to forage”.

Davis et al. (2013) showed that in arid Australian aquatic systems, perennial spring sites represent both ecological refuges for mobile taxa, and evolutionary refugia for species with low dispersal capabilities. Although this study focused on aquatic invertebrates, it presents a concept that could plausibly apply to vertebrates. Springs are likely to mitigate the impact of climate change on the dispersals of animals, hominins included, as they would provide potable water during dry periods, even in dry and very dry areas where many lakes are saline and rivers seasonal (Barboni, 2014; Cuthbert et al., 2017; Cuthbert and Ashley, 2014). In arid regions or during less favorable (more arid) climatic periods, groundwater springs may represent, for mobile taxa, ecological refuges or "stepping stones" between sites with more permanent water. Groundwater-fed sites in arid regions are also likely to contain relict and short-range endemic species as the habitats they provide is climatically decoupled (Davis et al., 2013; Harvey, 2002).

During arid climatic periods, spring sites could become sites of intense competition among species with overlapping ecological niches, therefore modifying the intensity of biotic interactions. At Olduvai, the paleosurface underlying Tuff IC dated ca 1.84 Ma ago (Deino, 2012), which includes evidence for springs and water holes bordered by palm groves (Arraiz, 2017; Ashley et al., 2010a; Dominguez-Rodrigo et al., 2010), has provided remains of three sympatric hominin species (*Paranthropus boisei*, *Homo habilis*, and likely *Homo erectus*) (Dominguez-Rodrigo et al., 2015). In this paleosurface were also recognized a carnivore kill site (AMK, Aramendi et al., 2017) and several hominin sites with evidence for carcass butchering such as FLK Zinj site and the newly discovered PTK and DS sites (Arráiz et al., 2017; Dominguez-Rodrigo and Cobo-Sánchez, 2017). These new finds are further evidence that, in Olduvai paleolake basin, freshwater springs were essential in the landscape for both hominins and other elements of the fauna. A review of lithic and faunal assemblages through Olduvai Bed I and Bed II (1.89 – 1.3 Ma) showed no correlation between

- 59–81. <https://doi.org/10.1086/674530>
- Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Ashley, G.M., Diez-Martin, F., Barboni, D., Prendergast, M.E., Yravedra, J., Barba, R., Sanchez, A., Baquedano, E., Pickering, T.R., 2010. New excavations at the FLK Zinjanthropus site and its surrounding landscape and their behavioral implications. *Quat. Res.* 74, 315–332. <https://doi.org/10.1016/j.yqres.2010.07.003>
- Domínguez-Rodrigo, M., Cobo-Sánchez, L., 2017. A spatial analysis of stone tools and fossil bones at FLK Jinj 22 and PTK I (Bed I, Olduvai Gorge, Tanzania) and its bearing on the social organization of early humans. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* <https://doi.org/10.1016/j.palaeo.2017.04.010>
- Dominguez-Rodrigo, M., Lopez-Saez, J.A., Vincens, A., Alcalá, L., Luque, L., Serrallonga, J., 2001. Fossil pollen from the Upper Humbu Formation of Peninj (Tanzania): hominid adaptation to a dry open Plio-Pleistocene savanna environment. *J. Hum. Evol.* 40, 151–157. <https://doi.org/10.1006/jhev.2000.0440>
- Domínguez-Rodrigo, M., Musiba, C.M., 2010. How Accurate are Paleoecological Reconstructions of Early Paleontological and Archaeological Sites? *Evol. Biol.* 37, 128–140. <https://doi.org/10.1007/s11692-010-9087-2>
- Domínguez-Rodrigo, M., Pickering, T.R., Almécija, S., Heaton, J.L., Baquedano, E., Mabulla, A., Uribelarrea, D., 2015. Earliest modern human-like hand bone from a new >1.84-million-year-old site at Olduvai in Tanzania. *Nat. Commun.* 6, 7987. <https://doi.org/10.1038/ncomms8987>
- Doran, T.L., Herries, A.I., Hopley, P.J., Sombroek, H., Hellstrom, J., Hodge, E., Kuhn, B.F., 2015. Assessing the paleoenvironmental potential of Pliocene to Holocene tufa deposits along the Ghaap Plateau escarpment (South Africa) using stable isotopes. *Quat. Res.* 84, 133–143.
- Du, A., Robinson, J.R., Rowan, J., Lazagabaster, I.A., Behrensmeyer, A.K., 2019. Stable carbon isotopes from paleosol carbonate and herbivore enamel document differing paleovegetation signals in the eastern African Plio-Pleistocene. *Rev. Palaeobot. Palynol.* 261, 41–52. <https://doi.org/10.1016/j.revpalbo.2018.11.003>
- Dunbar, R. and P., 1975. Guereza monkeys will they become extinct in Ethiopia? *Walia* 1975, 14–15.
- Dunbar, R.I.M., 1987. Habitat quality, population dynamics, and group composition in *Colobus* Monkeys (*Colobus guereza*). *Int. J. Primatol.* 8, 299–329. <https://doi.org/10.1007/BF02737386>
- Dunbar, R.I.M., Dunbar, E.P., 1974. Ecology and Population Dynamics of *Colobus guereza* in Ethiopia. *Folia Primatol. (Basel)* 21, 188–208. <https://doi.org/10.1159/000155600>
- Edwards, S., Demissew, S., Hedberg, I., 1997. Flora of Ethiopia and Eritrea: volume 6. Hydrocharitaceae to Arecaceae. Addis Ababa Natl. Herb. Biol. Dep. Addis Ababa Univ.
- Egeland, C.P., 2014. Taphonomic estimates of competition and the role of carnivore avoidance in hominin site use within the Early Pleistocene Olduvai Basin. *Quat. Int.*, The Evolution of Hominin Behavior during the Oldowan-Acheulian Transition: Recent Evidence from Olduvai Gorge and Peninj (Tanzania) 322–323, 95–106. <https://doi.org/10.1016/j.quaint.2013.11.021>
- Feakins, S.J., deMenocal, P.B., 2010. Global and African regional climate during the Cenozoic, in: Werdelin, L. (Ed.), *Cenozoic Mammals of Africa*. University of California Press, pp. 45–56.
- Feakins, S.J., Levin, N.E., Liddy, H.M., Sieracki, A., Eglinton, T.I., Bonnefille, R., 2013. Northeast African vegetation

2009. Hominin use of springs and wetlands: Paleoclimate and archaeological records from Olduvai Gorge (similar to 1.79-1.74 Ma). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 272, 1–16. <https://doi.org/10.1016/j.palaeo.2008.10.016>
- Bailey, G.N., Reynolds, S.C., King, G.C.P., 2011. Landscapes of human evolution: models and methods of tectonic geomorphology and the reconstruction of hominin landscapes. *J. Hum. Evol.* 60, 257–280. <https://doi.org/10.1016/j.jhevol.2010.01.004>
- Bamford, M.K., Albert, R.M., Cabanes, D., 2006. Plio-Pleistocene macroplant fossil remains and phytoliths from Lowermost Bed II in the eastern palaeolake margin of Olduvai Gorge, Tanzania. *Quat. Int.* 148, 95–112. <https://doi.org/10.1016/j.quaint.2005.11.027>
- Barboni, D., 2014. Vegetation of Northern Tanzania during the Plio-Pleistocene: A synthesis of the paleobotanical evidences from Laetoli, Olduvai, and Peninj hominin sites. *Quat. Int.*, The Evolution of Hominin Behavior during the Oldowan-Acheulian Transition: Recent Evidence from Olduvai Gorge and Peninj (Tanzania) 322–323, 264–276. <https://doi.org/10.1016/j.quaint.2014.01.016>
- Barboni, D., Ashley, G.M., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Baquedano, E., 2010. Phytoliths infer locally dense and heterogeneous paleovegetation at FLK North and surrounding localities during upper Bed I time, Olduvai Gorge, Tanzania. *Quat. Res.* 74, 344–354. <https://doi.org/10.1016/j.yqres.2010.09.005>
- Barboni, D., Bremond, L., 2009. Phytoliths of East African grasses: An assessment of their environmental and taxonomic significance based on floristic data. *Rev. Palaeobot. Palynol.* 158, 29–41. <https://doi.org/10.1016/j.revpalbo.2009.07.002>
- Barboni, D., Bremond, L., Bonnefille, R., 2007. Comparative study of modern phytolith assemblages from inter-tropical Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 246, 454–470. <https://doi.org/10.1016/j.palaeo.2006.10.012>
- Bender, R., Tobias, P.V., Bender, N., 2012. The savannah hypotheses: origin, reception and impact on paleoanthropology. *Hist. Philos. Life Sci.* 147–184.
- Benito-Calvo, A., Barfod, D., McHenry, L.J., De la Torre, I., 2014. The geology and chronology of the Acheulean deposits in the Mieso area (East-Central Ethiopia). *J. Hum. Evol.* 76, 26–38.
- Beverly, E.J., Driese, S.G., Peppe, D.J., Johnson, C.R., Michel, L.A., Faith, J.T., Tryon, C.A., Sharp, W.D., 2015. Recurrent spring-fed rivers in a Middle to Late Pleistocene semi-arid grassland: Implications for environments of early humans in the Lake Victoria Basin, Kenya. *Sedimentology* 62, 1611–1635.
- Bogani, M.T., Noriega-Felix, N., Vidal-Aguilar, S.L., Findley, L.T., Lytle, D.A., Gutiérrez-Ruacho, O.G., Alvarado-Castro, J.A., Varela-Romero, A., 2014. Biogeography and conservation of aquatic fauna in spring-fed tropical canyons of the southern Sonoran Desert, Mexico. *Biodivers. Conserv.* 23, 2705–2748.
- Bonnefille, R., 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Glob. Planet. Change* 72, 390–411. <https://doi.org/10.1016/j.gloplacha.2010.01.015>
- Bonnefille, R., Anupama, K., Barboni, D., Pascal, J., Prasad, S., Sutra, J.P., 1999. Modern pollen spectra from tropical South India and Sri Lanka: altitudinal distribution. *J. Biogeogr.* 26, 1255–1280. <https://doi.org/10.1046/j.1365-2699.1999.00359.x>

- Bonnefille, R., Riollet, G., 1987. Palynological spectra from the Upper Laetoli Beds, in: Leakey, M.D., Harris, J.M. (Eds.), *The Pliocene Site of Laetoli, Northern Tanzania*. Oxford, pp. 52–61.
- Box, E.E., 1981. Macroclimate and plant forms: an introduction to predictive modeling in phytogeography, Tasks for vegetation science. Junk, The Hague.
- Bremond, L., Alexandre, A., Hely, C., Guiot, J., 2005. A phytolith index as a proxy of tree cover density in tropical areas: Calibration with Leaf Area Index along a forest-savanna transect in southeastern Cameroon. *Glob. Planet. Change* 45, 277–293. <https://doi.org/10.1016/j.gloplacha.2004.09.002>
- Bretzler, A., Osenbrück, K., Gloaguen, R., Ruprecht, J.S., Kebede, S., Stadler, S., 2011. Groundwater origin and flow dynamics in active rift systems – A multi-isotope approach in the Main Ethiopian Rift. *J. Hydrol.* 402, 274–289. <https://doi.org/10.1016/j.jhydrol.2011.03.022>
- Brown, M.E., de Beurs, K., Vrieling, A., 2010. The response of African land surface phenology to large scale climate oscillations. *Remote Sens. Environ.* 114, 2286–2296. <https://doi.org/10.1016/j.rse.2010.05.005>
- Budd, D.A., Pack, S.M., Fogel, M.L., 2002. The destruction of paleoclimatic isotopic signals in Pleistocene carbonate soil nodules of Western Australia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 188, 249–273. [https://doi.org/10.1016/S0031-0182\(02\)00588-6](https://doi.org/10.1016/S0031-0182(02)00588-6)
- Butzer, K.W., 1973. Spring Sediments from the Acheulian Site of Amanzi (Uitenhage District, South Africa) 299–319.
- Cabanes, D., Shahack-Gross, R., 2015. Understanding fossil phytolith preservation: the role of partial dissolution in paleoecology and archaeology. *PloS One* 10, e0125532.
- Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I., Cox, E.J., 2012. Crenic habitats, hotspots for freshwater biodiversity conservation: toward an understanding of their ecology. *Freshw. Sci.* 31, 463–480. <https://doi.org/10.1899/11-111.1>
- Cantonati, M., Segadelli, S., Ogata, K., Tran, H., Sanders, D., Gerecke, R., Rott, E., Filippini, M., Gargini, A., Celico, F., 2016. A global review on ambient Limestone-Precipitating Springs (LPS): Hydrogeological setting, ecology, and conservation. *Sci. Total Environ.* 568, 624–637.
- Carr, C.J., 1998. Patterns of vegetation along the Omo River in southwest Ethiopia. *Plant Ecol.* 135, 135–163.
- Caylor, K.K., Shugart, H.H., 2006. Pattern and process in savanna ecosystems, in: D'Odorico, P., Porporato, A. (Eds.), *Dryland Ecohydrology*. Springer, The Netherlands, pp. 259–281.
- Cerling, T.E., Brown, F.H., Wynn, J.G., 2014. On the environment of Aramis: A comment on White in Domínguez-Rodrigo. *Curr. Anthropol.* 55, 469–470. <https://doi.org/10.1086/677210>
- Cerling, T.E., Levin, N.E., Quade, J., Wynn, J.G., Fox, D.L., Kingston, J.D., Klein, R.G., Brown, F.H., 2010. Comment on the paleoenvironment of *Ardipithecus ramidus*. *SCIENCE* 328. <https://doi.org/10.1126/science.1185274>
- Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., Mace, W., Macharia, A.N., Quade, J., Remien, C.H., 2011. Woody cover and hominin environments in the past 6 million years. *Nature* 476, 51–56. <https://doi.org/10.1038/nature10306>
- Chorowicz, J., 2005. The East African rift system. *J. Afr. Earth Sci.* 43, 379–410. <https://doi.org/10.1016/j.jafrearsci.2005.07.019>
- Churcher, C.S., Kleindienst, M.R., Schwarcz, H.P., 1999. Faunal remains from a Middle Pleistocene lacustrine marl in Dakhleh Oasis, Egypt:

hominin sites and predator risk, and suggests instead, that water and tree cover “played more proximal roles” for Early Pleistocene hominins than carnivore avoidance (Egeland, 2014). Hence, although freshwater would attract both prey and predators, and would become places of intensified predation risk, the presence of trees likely offered a crucial advantage of safety for primates, including early hominins.

#### 4.4. Perennial water at the continental scale: springs versus permanent rivers

We have found 50 hominin-spring associations and just 12 that include localities dated to >1 Ma (Fig.1, Table 1). Although it is likely that spring deposits may have been overlooked in some sites, this number is low compared with e.g. the ~110 hominin-bearing localities of the Plio-Pleistocene Shungura Formation in the Turkana Basin that do not seem to display evidence for groundwater-fed habitats (JR Boisserie, *pers. com*). Although the large hominin occurrence is the Omo Valley points to the importance of surface water and riparian micro-habitats rather than groundwater in this region, what it really means is that permanent water and the wooded micro-habitats it may sustain are the key parameters to the ecology and probably the dispersal of early hominins.

Perennial rivers on the one hand, and springs on the other hand both contribute permanent water and sustain wooded habitats, but at different temporal and spatial scales. The Shungura Formation is a continuous record of fluvial, deltaic, and lacustrine environments in the Omo Valley dating since the Pliocene to the mid-Pleistocene indicating that the river system was active even during driest periods (de Heinzelin, 1983; McDougall et al., 2012). It is likely that the Omo River has been active since relief was created, i.e. since the Ethiopian volcanic doming >23 million years ago (Rooney, 2017). Same reasoning applies to the Awash River. Permanent rivers, hence, are likely to provide highly stable water resource and wooded micro-habitats on much longer timescales ( $10^6$  –  $10^7$  years) than springs, which start or cease being active according to much more ‘frequent’ intra-rift faulting and basin formation ( $10^3$  –  $10^6$  years).

Permanent rivers, in addition, may represent hundreds of km-long wooded

corridors that could potentially connect different valleys through the highlands, making inter-regional dispersal possible for faunas favoring wooded habitats (and relatively tolerant to lower temperatures). Landscape heterogeneity (due to topography, volcanism, tectonism) in addition to the micro-habitat diversity created by the hydrographic network in the rift valleys break off the apparent homogeneity of the savanna biome at the regional scale (Bailey et al., 2011; Reynolds et al., 2015, 2011). Patterns of faunal diversity and their relationship with the heterogeneity of landscapes and habitats in various valleys during the Pliocene would be worth investigating.

#### 5. Conclusions

The geological context, the present-day vegetation, and the pollen and phytolith signatures of several modern springs from two regions in Eastern Africa that have numerous paleoanthropological sites, namely the Awash Valley (Ethiopia) and the Crater Highlands Region (Tanzania) were sampled, generating new plant microfossil datasets of 20 pollen and 19 phytolith samples. The new phytolith dataset that we have combined with previously published data (Albert et al., 2015; Arraiz, 2017; Barboni et al., 2007; Novello, 2012; Novello et al., 2017) now makes up a dataset of 265 surface samples, which allows a more accurate interpretation of ancient spring and wetland environments in arid Africa.

Groundwater-associated forests, palm groves, and wetlands, despite being small natural features, play essential roles at the species and ecosystem levels, particularly in arid and sub-arid regions. Springs and groundwater-fed areas are spatially localized, but temporally persistent features that are independent of the regional climate drivers. They contribute to additional soil moisture that generates the development of species-rich and structurally complex microhabitats in arid regions.

Interpretation of the paleoecology suggests that in the East African Rift where low, highly seasonal rainfall and high evapo-transpiration limit plant growth in many areas, springs and groundwater-fed zones play a major role in ecosystem functioning. Springs are commonly localized features (oases) within a landscape;

they are limited in space, but yield a dependable daily supply of water and provide ecological continuity through time. At the local scale, springs create a microclimate, distinctive vegetation, increase soil nutrients, species richness, structural complexity, and provide habitat for animals. At the landscape scale, they represent hydro-refugia favoring increased connectivity among animals and allowing migrations during dry periods. Hence, the palm and grass-rich spring woodland that made up the habitat of *Ardipithecus ramidus* likely played a crucial role on the feeding ecology, locomotion, and evolutionary trend of this early hominin in northeastern Ethiopia and, potentially of other hominins elsewhere in East Africa.

Aramis paleoenvironment as inferred from the sampling of 9 km-long west-east transect in the Middle Awash Valley is interpreted as a mosaic of micro-habitats including forest patches and palm groves closely associated with the presence of groundwater, and open grasslands. Aramis wooded micro-habitats represent by no means the paleoenvironment at the landscape or regional scale, but localized, azonal patches of wooded vegetation within the much larger, climatically-driven, savanna biome.

More generally, our study shows that it is permanent water and the wooded micro-habitats that it may sustain that are likely the key parameters to the ecology and dispersal of early hominins. Hominin and faunal evolution in general was likely driven by a complex interplay between abiotic factors (tectonics, hydrography, climate) and biotic interactions at various spatial and temporal scales, and not simply by climatically-driven biome changes.

#### Research data

SOM 1. Site data and pollen counts for samples from the Awash Valley (Ethiopia) (samples BB17-) and from the region of Lake Eyasi - Lake Manyara (Tanzania) (samples DB11-).  
<http://dx.doi.org/10.17632/dsgd78tfjz.1 - file-12c6e371-ea9a-4502-87ce-7096c7f6214e>

SOM 2. Site data and raw phytolith counts for 265 surface soil samples and for the 41 fossil samples from Aramis Member, Sagantole Formation.

<http://dx.doi.org/10.17632/myg68hsjvf.1>

SOM 3. Site data and raw phytolith counts for the 41 fossil samples from Aramis Member, Sagantole Formation (partially published in WoldeGabriel et al., 2009).

<http://dx.doi.org/10.17632/4crtycn7rk.1>

#### Acknowledgments

For financial support during fieldwork we are grateful to The Olduvai Paleontological and Paleoenvironment Project (TOPPP, M. Dominguez-Rodrigo, E. Baquedano, A.Z.P. Mabulla), the NSF (grant number 0321893 HOMINID-RHOI), National Geographic Waitt grant (#W168-11 to GM Ashley) and CEREGE (APIC 2017). We are grateful to J-R. Boissarie, M. Dominguez-Rodrigo, T. White, and S. Ambrose for their insightful comments to early drafts of this manuscript, to Melaku Wondafrash, Admassu Addi, Amare, Elema, and Abel Belay for help during fieldwork in Ethiopia. We are grateful to H. Bunn, Chris and Nani Schmeling at Kisima Ngeda, Shabani, Joseph Masoy, Sikwasi and the Hadzabe for their help during fieldwork in Tanzania. We are grateful to K. Koops (Cambridge, UK) for providing surface samples from Guinea, to A. Novello and R-M. Albert for sharing their data from Chad and Tanzania, respectively, and to A. Galy (CRPG, France) for discussing the issue of carbon isotopes in pedogenic carbonate nodules. Guillaume Buchet helped with pollen counting. We are deeply saddened by his passing.

#### References

- Albert, R.M., Bamford, M.K., 2012. Vegetation during UMBI and deposition of Tuff IF at Olduvai Gorge, Tanzania (ca. 1.8 Ma) based on phytoliths and plant remains. *J. Hum. Evol.* 63, 342–350. <https://doi.org/10.1016/j.jhevol.2011.05.010>
- Albert, R.M., Bamford, M.K., Esteban, I., 2015. Reconstruction of ancient palm vegetation landscapes using a phytolith approach. *Quat. Int.* 369, 51–66.
- Albert, R.M., Bamford, M.K., Stanistreet, I.G., Stollhofen, H., Rivera-Rondón, C.A., Njau, J.K., Blumenschine, R.J., 2018. River-fed wetland palaeo-vegetation and palaeoecology at the HWK W site,

- Bed I, Olduvai Gorge. Rev. Palaeobot. Palynol. <https://doi.org/10.1016/j.revpalbo.2018.09.010>
- Aramendi, J., Uribelarrea, D., Arriaza, M.C., Arráiz, H., Barboni, D., Yravedra, J., Ortega, M.C., Gidna, A., Mabulla, A., Baquedano, E., others, 2017. The paleoecology and taphonomy of AMK (Bed I, Olduvai Gorge) and its contributions to the understanding of the “Zinj” paleolandscape. Palaeogeogr. Palaeoclimatol. Palaeoecol. 488, 35–49.
- Arraiz, H., 2017. Paleovegetation and stone tool use at a selection of hominin sites and their associated landscapes from Olduvai Gorge (Tanzania): a study of plant microfossils (Doctorat). Universidad Complutense De Madrid, Madrid, Spain.
- Arráiz, H., Barboni, D., Ashley, G.M., Mabulla, A., Baquedano, E., Domínguez-Rodrigo, M., 2017. The FLK Zinj paleolandscape: Reconstruction of a 1.84 Ma wooded habitat in the FLK Zinj-AMK-PTK-DS archaeological complex, Middle Bed I (Olduvai Gorge, Tanzania). Palaeogeogr. Palaeoclimatol. Palaeoecol. <https://doi.org/10.1016/j.palaeo.2017.04.025>
- Ash, J., Atkins, J., 2010. Birds of Ethiopia and Eritrea: An Atlas of Distribution. Bloomsbury Publishing.
- Ashley, G.M., Barboni, D., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Diez-Martin, F., Barba, R., Baquedano, E., 2010a. A spring and wooded habitat at FLK Zinj and their relevance to origins of human behavior. Quat. Res. 74, 304–314. <https://doi.org/10.1016/j.yqres.2010.07.015>
- Ashley, G.M., Barboni, D., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Diez-Martin, F., Barba, R., Baquedano, E., 2010b. Paleoenvironmental and paleoecological reconstruction of a freshwater oasis in savannah grassland at FLK North, Olduvai Gorge, Tanzania. Quat. Res. 74, 333–343. <https://doi.org/10.1016/j.yqres.2010.08.006>
- Ashley, G.M., Bunn, H.T., Delaney, J.S., Barboni, D., Dominguez-Rodrigo, M., Mabulla, A.Z.P., Gurto, A.N., Baluyot, R., Beverly, E.J., Baquedano, E., 2014a. Paleoclimatic and paleoenvironmental framework of FLK North archaeological site, Olduvai Gorge, Tanzania. Quat. Int. 322–323, 54–65. <https://doi.org/10.1016/j.quaint.2013.08.052>
- Ashley, G.M., de Wet, C.B., Barboni, D., Magill, C.R., 2016. Subtle signatures of seeps: Record of groundwater in a Dryland, DK, Olduvai Gorge, Tanzania. Depositional Rec. 2, 4–21. <https://doi.org/10.1002/dep2.11>
- Ashley, G.M., De Wet, C.B., Dominguez-Rodrigo, M., Karis, A.M., O'Reilly, T., Baluyot, R., 2014b. Freshwater limestone in an arid rift basin: a goldilocks effect. J. Sediment. Res. 84, 988–1004.
- Ashley, G.M., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Baquedano, E., 2010c. Sedimentary geology and human origins: A fresh look at Olduvai Gorge, Tanzania. J. Sediment. Res. 80, 703–709. <https://doi.org/10.2110/jsr.2010.066>
- Ashley, G.M., Goman, M., Hover, V., Owen, R., Renaut, R., Muasya, A., 2002. Artesian blister wetlands, a perennial water resource in the semi-arid rift valley of East Africa. Wetlands 22, 686–695. [https://doi.org/10.1672/0277-5212\(2002\)022\[0686:ABWAPW\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2002)022[0686:ABWAPW]2.0.CO;2)
- Ashley, G.M., Mworia, J., Muasya, A., Owen, R., Driese, S., Hover, V., Renaut, R., Goman, M., Mathai, S., Blatt, S., 2004. Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Lobi Swamp, Kenya, East Africa. Sedimentology 51, 1301–1321. <https://doi.org/10.1111/j.1365-3091.2004.00671.x>
- Ashley, G.M., Tactikos, J.C., Owen, R.B.,

2009. Hominin use of springs and wetlands: Paleoclimate and archaeological records from Olduvai Gorge (similar to 1.79–1.74 Ma). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 272, 1–16. <https://doi.org/10.1016/j.palaeo.2008.10.016>
- Bailey, G.N., Reynolds, S.C., King, G.C.P., 2011. Landscapes of human evolution: models and methods of tectonic geomorphology and the reconstruction of hominin landscapes. *J. Hum. Evol.* 60, 257–280. <https://doi.org/10.1016/j.jhevol.2010.01.004>
- Bamford, M.K., Albert, R.M., Cabanes, D., 2006. Plio-Pleistocene macroplant fossil remains and phytoliths from Lowermost Bed II in the eastern palaeolake margin of Olduvai Gorge, Tanzania. *Quat. Int.* 148, 95–112. <https://doi.org/10.1016/j.quaint.2005.11.027>
- Barboni, D., 2014. Vegetation of Northern Tanzania during the Plio-Pleistocene: A synthesis of the paleobotanical evidences from Laetoli, Olduvai, and Peninj hominin sites. *Quat. Int., The Evolution of Hominin Behavior during the Oldowan-Acheulian Transition: Recent Evidence from Olduvai Gorge and Peninj (Tanzania)* 322–323, 264–276. <https://doi.org/10.1016/j.quaint.2014.01.016>
- Barboni, D., Ashley, G.M., Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Baquedano, E., 2010. Phytoliths infer locally dense and heterogeneous paleovegetation at FLK North and surrounding localities during upper Bed I time, Olduvai Gorge, Tanzania. *Quat. Res.* 74, 344–354. <https://doi.org/10.1016/j.yqres.2010.09.005>
- Barboni, D., Bremond, L., 2009. Phytoliths of East African grasses: An assessment of their environmental and taxonomic significance based on floristic data. *Rev. Palaeobot. Palynol.* 158, 29–41. <https://doi.org/10.1016/j.revpalbo.2009.07.002>
- Barboni, D., Bremond, L., Bonnefille, R., 2007. Comparative study of modern phytolith assemblages from inter-tropical Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 246, 454–470. <https://doi.org/10.1016/j.palaeo.2006.10.012>
- Bender, R., Tobias, P.V., Bender, N., 2012. The savannah hypotheses: origin, reception and impact on paleoanthropology. *Hist. Philos. Life Sci.* 147–184.
- Benito-Calvo, A., Barfod, D., McHenry, L.J., De la Torre, I., 2014. The geology and chronology of the Acheulean deposits in the Mieso area (East-Central Ethiopia). *J. Hum. Evol.* 76, 26–38.
- Beverly, E.J., Driese, S.G., Peppe, D.J., Johnson, C.R., Michel, L.A., Faith, J.T., Tryon, C.A., Sharp, W.D., 2015. Recurrent spring-fed rivers in a Middle to Late Pleistocene semi-arid grassland: Implications for environments of early humans in the Lake Victoria Basin, Kenya. *Sedimentology* 62, 1611–1635.
- Bogdan, M.T., Noriega-Felix, N., Vidal-Aguilar, S.L., Findley, L.T., Lytle, D.A., Gutiérrez-Ruacho, O.G., Alvarado-Castro, J.A., Varela-Romero, A., 2014. Biogeography and conservation of aquatic fauna in spring-fed tropical canyons of the southern Sonoran Desert, Mexico. *Biodivers. Conserv.* 23, 2705–2748.
- Bonnefille, R., 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Glob. Planet. Change* 72, 390–411. <https://doi.org/10.1016/j.gloplacha.2010.01.015>
- Bonnefille, R., Anupama, K., Barboni, D., Pascal, J., Prasad, S., Sutra, J.P., 1999. Modern pollen spectra from tropical South India and Sri Lanka: altitudinal distribution. *J. Biogeogr.* 26, 1255–1280. <https://doi.org/10.1046/j.1365-2699.1999.00359.x>

- Bonnefille, R., Riollet, G., 1987. Palynological spectra from the Upper Laetolil Beds, in: Leakey, M.D., Harris, J.M. (Eds.), The Pliocene Site of Laetoli, Northern Tanzania. Oxford, pp. 52–61.
- Box, E.E., 1981. Macroclimate and plant forms: an introduction to predictive modeling in phytogeography, Tasks for vegetation science. Junk, The Hague.
- Bremond, L., Alexandre, A., Hely, C., Guiot, J., 2005. A phytolith index as a proxy of tree cover density in tropical areas: Calibration with Leaf Area Index along a forest-savanna transect in southeastern Cameroon. *Glob. Planet. Change* 45, 277–293. <https://doi.org/10.1016/j.gloplacha.2004.09.002>
- Bretzler, A., Osenbrück, K., Gloaguen, R., Ruprecht, J.S., Kebede, S., Stadler, S., 2011. Groundwater origin and flow dynamics in active rift systems – A multi-isotope approach in the Main Ethiopian Rift. *J. Hydrol.* 402, 274–289. <https://doi.org/10.1016/j.jhydrol.2011.03.022>
- Brown, M.E., de Beurs, K., Vrieling, A., 2010. The response of African land surface phenology to large scale climate oscillations. *Remote Sens. Environ.* 114, 2286–2296. <https://doi.org/10.1016/j.rse.2010.05.005>
- Budd, D.A., Pack, S.M., Fogel, M.L., 2002. The destruction of paleoclimatic isotopic signals in Pleistocene carbonate soil nodules of Western Australia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 188, 249–273. [https://doi.org/10.1016/S0031-0182\(02\)00588-6](https://doi.org/10.1016/S0031-0182(02)00588-6)
- Butzer, K.W., 1973. Spring Sediments from the Acheulian Site of Amanzi (Uitenhage District, South Africa) 299–319.
- Cabanes, D., Shahack-Gross, R., 2015. Understanding fossil phytolith preservation: the role of partial dissolution in paleoecology and archaeology. *PloS One* 10, e0125532.
- Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I., Cox, E.J., 2012. Crenic habitats, hotspots for freshwater biodiversity conservation: toward an understanding of their ecology. *Freshw. Sci.* 31, 463–480. <https://doi.org/10.1899/11-111.1>
- Cantonati, M., Segadelli, S., Ogata, K., Tran, H., Sanders, D., Gerecke, R., Rott, E., Filippini, M., Gargini, A., Celico, F., 2016. A global review on ambient Limestone-Precipitating Springs (LPS): Hydrogeological setting, ecology, and conservation. *Sci. Total Environ.* 568, 624–637.
- Carr, C.J., 1998. Patterns of vegetation along the Omo River in southwest Ethiopia. *Plant Ecol.* 135, 135–163.
- Caylor, K.K., Shugart, H.H., 2006. Pattern and process in savanna ecosystems, in: D'Odorico, P., Porporato, A. (Eds.), *Dryland Ecohydrology*. Springer, The Netherlands, pp. 259–281.
- Cerling, T.E., Brown, F.H., Wynn, J.G., 2014. On the environment of Aramis: A comment on White in Domínguez-Rodrigo. *Curr. Anthropol.* 55, 469–470. <https://doi.org/10.1086/677210>
- Cerling, T.E., Levin, N.E., Quade, J., Wynn, J.G., Fox, D.L., Kingston, J.D., Klein, R.G., Brown, F.H., 2010. Comment on the paleoenvironment of *Ardipithecus ramidus*. *SCIENCE* 328. <https://doi.org/10.1126/science.1185274>
- Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., Mace, W., Macharia, A.N., Quade, J., Remien, C.H., 2011. Woody cover and hominin environments in the past 6 million years. *Nature* 476, 51–56. <https://doi.org/10.1038/nature10306>
- Chorowicz, J., 2005. The East African rift system. *J. Afr. Earth Sci.* 43, 379–410. <https://doi.org/10.1016/j.jafrearsci.2005.07.019>
- Churcher, C.S., Kleindienst, M.R., Schwarcz, H.P., 1999. Faunal remains from a Middle Pleistocene lacustrine marl in Dakhleh Oasis, Egypt:

- palaeoenvironmental reconstructions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 154, 301–312. [https://doi.org/10.1016/S0031-0182\(99\)00104-2](https://doi.org/10.1016/S0031-0182(99)00104-2)
- Collier, K.J., Smith, B.J., 2006. Distinctive invertebrate assemblages in rockface seepages enhance lotic biodiversity in northern New Zealand, in: Marine, Freshwater, and Wetlands Biodiversity Conservation. Springer, pp. 223–248.
- Collura, L.V., Neumann, K., 2017. Wood and bark phytoliths of West African woody plants. *Quat. Int.*, 9th International Meeting of Phytolith Research (IMPR) 434, 142–159. <https://doi.org/10.1016/j.quaint.2015.12.070>
- Contoux, C., Jost, A., Ramstein, G., Sepulchre, P., Krinner, G., Schuster, M., 2013. Megalake Chad impact on climate and vegetation during the late Pliocene and the mid-Holocene. *Clim. Past* 9, 1417–1430.
- Cuthbert, M.O., Ashley, G.M., 2014. A spring forward for hominin evolution in East Africa. *PLOS ONE* 9, e107358. <https://doi.org/10.1371/journal.pone.0107358>
- Cuthbert, M.O., Gleeson, T., Reynolds, S.C., Bennett, M.R., Newton, A.C., McCormack, C.J., Ashley, G.M., 2017. Modelling the role of groundwater hydro-refugia in East African hominin evolution and dispersal. *Nat. Commun.* 8, 15696. <https://doi.org/10.1038/ncomms15696>
- Dachy, T., Bröis, F., Marchand, S., Minotti, M., Lesur, J., Wuttmann, M., 2018. Living in an Egyptian Oasis: Reconstruction of the Holocene archaeological sequence in Kharga. *Afr. Archaeol. Rev.* 1–36.
- Dart, R.A., 1925. *Australopithecus africanus*: the man-ape of South Africa. *Nature* 115, 195–199.
- Davis, J., Pavlova, A., Thompson, R., Sunnucks, P., 2013. Evolutionary refugia and ecological refuges: key concepts for conserving Australian arid zone freshwater biodiversity under climate change. *Glob. Change Biol.* 19, 1970–1984. <https://doi.org/10.1111/gcb.12203>
- Davis, J.A., Kerezsy, A., Nicol, S., 2017. Springs: Conserving perennial water is critical in arid landscapes. *Biol. Conserv.*, Small Natural Features 211, 30–35. <https://doi.org/10.1016/j.biocon.2016.12.036>
- de Heinzelin, J. (Ed.), 1983. The Omo Group, Musée Royal de l'Afrique Centrale, Annales, Série in 8: Sciences Géologiques.
- DeMenocal, P.B., 1995. Plio-Pleistocene African climate. *Science* 270, 53–59.
- Deocampo, D., Blumenshine, R., Ashley, G.M., 2002. Wetland diagenesis and traces of early hominids, Olduvai Gorge, Tanzania. *Quat. Res.* 57, 271–281. <https://doi.org/10.1006/qres.2001.2317>
- Deocampo, D.M., Tactikos, J.C., 2010. Geochemical gradients and artifact mass densities on the lowermost Bed II eastern lake margin (similar to 1.8 Ma), Olduvai Gorge, Tanzania. *Quat. Res.* 74, 411–423. <https://doi.org/10.1016/j.yqres.2010.09.004>
- Ditchfield, P., Harrison, T., 2011. Sedimentology, lithostratigraphy and depositional history of the Laetoli area, in: Paleontology and Geology of Laetoli: Human Evolution in Context. Springer, pp. 47–76.
- Djamali, M., Gondet, S., Ashjari, J., Aubert, C., Brisset, E., Longerey, J., Marriner, N., Mashkour, M., Miller, N.F., Naderi-Beni, A., Pourkerman, M., Rashidian, E., Rigot, J.-B., Shidrang, S., Thiéry, A., Gandouin, E., 2018. Karstic spring wetlands of the Persepolis Basin, southwest Iran: unique sediment archives of Holocene environmental change and human impacts. *Can. J. Earth Sci.* 55, 1158–1172. <https://doi.org/10.1139/cjes-2018-0065>
- Domínguez-Rodrigo, M., 2014. Is the “Savanna Hypothesis” a Dead Concept for Explaining the Emergence of the Earliest Hominins? *Curr. Anthropol.* 55,

- 59–81. <https://doi.org/10.1086/674530>
- Dominguez-Rodrigo, M., Bunn, H.T., Mabulla, A.Z.P., Ashley, G.M., Diez-Martin, F., Barboni, D., Prendergast, M.E., Yravedra, J., Barba, R., Sanchez, A., Baquedano, E., Pickering, T.R., 2010. New excavations at the FLK Zinjanthropus site and its surrounding landscape and their behavioral implications. *Quat. Res.* 74, 315–332. <https://doi.org/10.1016/j.yqres.2010.07.003>
- Domínguez-Rodrigo, M., Cobo-Sánchez, L., 2017. A spatial analysis of stone tools and fossil bones at FLK Jinj 22 and PTK I (Bed I, Olduvai Gorge, Tanzania) and its bearing on the social organization of early humans. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* <https://doi.org/10.1016/j.palaeo.2017.04.010>
- Dominguez-Rodrigo, M., Lopez-Saez, J.A., Vincens, A., Alcalá, L., Luque, L., Serrallonga, J., 2001. Fossil pollen from the Upper Humbu Formation of Peninj (Tanzania): hominin adaptation to a dry open Plio-Pleistocene savanna environment. *J. Hum. Evol.* 40, 151–157. <https://doi.org/10.1006/jhev.2000.0440>
- Domínguez-Rodrigo, M., Musiba, C.M., 2010. How Accurate are Paleoecological Reconstructions of Early Paleontological and Archaeological Sites? *Evol. Biol.* 37, 128–140. <https://doi.org/10.1007/s11692-010-9087-2>
- Domínguez-Rodrigo, M., Pickering, T.R., Almécija, S., Heaton, J.L., Baquedano, E., Mabulla, A., Urielarrea, D., 2015. Earliest modern human-like hand bone from a new >1.84-million-year-old site at Olduvai in Tanzania. *Nat. Commun.* 6, 7987. <https://doi.org/10.1038/ncomms8987>
- Doran, T.L., Herries, A.I., Hopley, P.J., Sombroek, H., Hellstrom, J., Hodge, E., Kuhn, B.F., 2015. Assessing the paleoenvironmental potential of Pliocene to Holocene tufa deposits along the Ghaap Plateau escarpment (South Africa) using stable isotopes. *Quat. Res.* 84, 133–143.
- Du, A., Robinson, J.R., Rowan, J., Lazagabaster, I.A., Behrensmeyer, A.K., 2019. Stable carbon isotopes from paleosol carbonate and herbivore enamel document differing paleovegetation signals in the eastern African Plio-Pleistocene. *Rev. Palaeobot. Palynol.* 261, 41–52. <https://doi.org/10.1016/j.revpalbo.2018.11.003>
- Dunbar, R. and P., 1975. Guereza monkeys will they become extinct in Ethiopia? *Walia* 1975, 14–15.
- Dunbar, R.I.M., 1987. Habitat quality, population dynamics, and group composition in *Colobus* Monkeys (*Colobus guereza*). *Int. J. Primatol.* 8, 299–329. <https://doi.org/10.1007/BF02737386>
- Dunbar, R.I.M., Dunbar, E.P., 1974. Ecology and Population Dynamics of *Colobus guereza* in Ethiopia. *Folia Primatol. (Basel)* 21, 188–208. <https://doi.org/10.1159/000155600>
- Edwards, S., Demissew, S., Hedberg, I., 1997. Flora of Ethiopia and Eritrea: volume 6. Hydrocharitaceae to Arecaceae. Addis Ababa Natl. Herb. Biol. Dep. Addis Ababa Univ.
- Egeland, C.P., 2014. Taphonomic estimates of competition and the role of carnivore avoidance in hominin site use within the Early Pleistocene Olduvai Basin. *Quat. Int.*, The Evolution of Hominin Behavior during the Oldowan-Acheulian Transition: Recent Evidence from Olduvai Gorge and Peninj (Tanzania) 322–323, 95–106. <https://doi.org/10.1016/j.quaint.2013.11.021>
- Feakins, S.J., deMenocal, P.B., 2010. Global and African regional climate during the Cenozoic, in: Werdelin, L. (Ed.), *Cenozoic Mammals of Africa*. University of California Press, pp. 45–56.
- Feakins, S.J., Levin, N.E., Liddy, H.M., Sieracki, A., Eglington, T.I., Bonnefille, R., 2013. Northeast African vegetation

- change over 12 m.y. *Geology* G33845.1. <https://doi.org/10.1130/G33845.1>
- Fensham, R.J., Silcock, J.L., Kerezsy, A., Ponder, W., 2011. Four desert waters: Setting arid zone wetland conservation priorities through understanding patterns of endemism. *Biol. Conserv.* 144, 2459–2467. <https://doi.org/10.1016/j.biocon.2011.06.024>
- Foulds, F.W.F., Shuttleworth, A., Sinclair, A., Alsharekh, A.M., Ghamdi, S.A., Inglis, R.H., Bailey, G.N., 2017. A large handaxe from Wadi Dabsa and early hominin adaptations within the Arabian Peninsula. *Antiquity* 91, 1421–1434. <https://doi.org/10.15184/ajy.2017.153>
- Fraysse, F., Cantais, F., Pokrovsky, O.S., Schott, J., Meunier, J.-D., 2006. Aqueous reactivity of phytoliths and plant litter: physico-chemical constraints on terrestrial biogeochemical cycle of silicon. *J. Geochem. Explor.* 88, 202–205.
- Fraysse, F., Pokrovsky, O.S., Schott, J., Meunier, J.-D., 2009. Surface chemistry and reactivity of plant phytoliths in aqueous solutions. *Chem. Geol.* 258, 197–206.
- Gallagher, T.M., Sheldon, N.D., 2016. Combining soil water balance and clumped isotopes to understand the nature and timing of pedogenic carbonate formation. *Chem. Geol.* 435, 79–91. <https://doi.org/10.1016/j.chemgeo.2016.04.023>
- Garrett, K.R., 2017. Multi-proxy reconstruction of a~1.3 Ma freshwater wetland, Olduvai Gorge, Tanzania (MS Thesis). Rutgers University-Graduate School-New Brunswick.
- Good, S.P., Caylor, K.K., 2011. Climatological determinants of woody cover in Africa. *Proc. Natl. Acad. Sci.* 108, 4902–4907. <https://doi.org/10.1073/pnas.1013100108>
- Gossa, T., Sahle, Y., Negash, A., 2012. A reassessment of the Middle and Later Stone Age lithic assemblages from Aladi Springs, Southern Afar Rift, Ethiopia. *Azania Archaeol. Res. Afr.* 47, 210–222.
- Greenway, P.J., Vesey-Fitzgerald, D.F., 1969. Vegetation of Lake Manyara National Park. *J. Ecol.* 57, 127–149. <https://doi.org/10.2307/2258212>
- Haile-Selassie, Y., Deino, A., Saylor, B., Umer, M., Latimer, B., 2007. Preliminary geology and paleontology of new hominid-bearing Pliocene localities in the central Afar region of Ethiopia. *Anthropol. Sci.* 115, 215–222.
- Harvey, M.S., 2002. Short-range endemism amongst the Australian fauna: some examples from non-marine environments. *Invertebr. Syst.* 16, 555–570. <https://doi.org/10.1071/is02009>
- Hill, C.L., 2001. Geologic Contexts of the Acheulian (Middle Pleistocene) in the Eastern Sahara. *Geoarchaeology Int. J.* 16, 65–94.
- Hodson, M., White, P., Mead, A., Broadley, M., 2005. Phylogenetic variation in the silicon composition of plants. *Ann. Bot.* 96, 1027–1046. <https://doi.org/10.1093/aob/mci255>
- Hopley, P.J., Herries, A.I.R., Baker, S.E., Kuhn, B.F., Menter, C.G., 2013. Brief communication: beyond the South African cave paradigm--*Australopithecus africanus* from Plio-Pleistocene paleosol deposits at Taung. *Am. J. Phys. Anthropol.* 151, 316–324. <https://doi.org/10.1002/ajpa.22272>
- Hunter Jr, M.L., Acuña, V., Bauer, D.M., Bell, K.P., Calhoun, A.J., Felipe-Lucia, M.R., Fitzsimons, J.A., González, E., Kinnison, M., Lindenmayer, D., 2017. Conserving small natural features with large ecological roles: a synthetic overview. *Biol. Conserv.* 211, 88–95.
- Hussein, I., Afework, B., Dereje, Y., 2017. Population structure and feeding ecology of Guereza (*Colobus guereza*) in Borena-Sayint National Park, northern Ethiopia. *Int. J. Biodivers. Conserv.* 9, 323–333. <https://doi.org/10.5897/IJBC2017.1114>
- Inglis, R.H., Sinclair, A.G., Alsharekh, A.M., Barfod, D., Chang, H.C., Fanning, P.C., Al Othaibi, D.T., Robson, H.K.,

- Shuttleworth, A., Stone, A., Bailey, G.N., 2017. Preliminary Report on UK-Saudi 2017 Fieldwork At Wadi Dabsa, Asir Province, Saudi Arabia (Fieldwork report). University of York.
- Jagher, R., Elsuede, H., Le Tensorer, J.-M., 2015. El Kowm Oasis, human settlement in the Syrian Desert during the Pleistocene. *L'anthropologie* 119, 542–580.
- Johnson, C.R., Ashley, G.M., De Wet, C.B., Dvoretzky, R., Park, L., Hover, V.C., Owen, R.B., McBrearty, S., 2009. Tufa as a record of perennial fresh water in a semi-arid rift basin, Kapthurin Formation, Kenya. *Sedimentology* 56, 1115–1137.  
<https://doi.org/10.1111/j.1365-3091.2008.01022.x>
- Johnson, C.R., McBrearty, S., 2012. Archaeology of middle Pleistocene lacustrine and spring paleoenvironments in the Kapthurin Formation, Kenya. *J. Anthropol. Archaeol.* 31, 485–499.
- Jolly-Saad, M.-C., Bonnefille, R., 2012. Lower Pliocene Fossil Wood from the Middle Awash Valley, Ethiopia. *Palaeontogr. Abt. B* 43–73.
- Kappelman, J., Alçıçek, M.C., Kazancı, N., Schultz, M., Özkul, M., Şen, S., 2008. First *Homo erectus* from Turkey and implications for migrations into temperate Eurasia. *Am. J. Phys. Anthropol.* 135, 110–116.
- Kassambara, A., Mundt, F., 2017. Factoextra: extract and visualize the results of multivariate data analyses.
- Kaya, F., Bibi, F., Źliobaitė, I., Eronen, J.T., Hui, T., Fortelius, M., 2018. The rise and fall of the Old World savannah fauna and the origins of the African savannah biome. *Nat. Ecol. Evol.* 2, 241.
- Kempf, E., 2009. Patterns of Water Use in Primates. *Folia Primatol.* 80, 275–294.
- Kleindienst, M.R., Schwarcz, H.P., Nicoll, K., Churcher, C.S., Frizano, J., Giegengack, R., Wiseman, M.F., 2008. Water in the desert: first report on Uranium-series dating of Caton-Thompson's and Gardner's 'classic' Pleistocene sequence at Refuf Pass, Kharga Oasis, in: *The Oasis Papers 2. Proceedings of the Second International Conference of the Dakhleh Oasis Project*. Oxbow Books Oxford, pp. 25–54.
- Lamarck, J.-B. de, 1809. *Philosophie zoologique, ou exposition des considérations relatives à l'histoire naturelle des animaux*. Dentu.
- Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* 25, 1–18.
- Leakey, M.D., 1971. *Olduvai Gorge: Excavations in Bed I and II, 1960-63*. Cambridge University Press.
- Lebatard, A.-E., Alçıçek, M.C., Rochette, P., Khatib, S., Vialet, A., Boulbes, N., Bourlès, D.L., Demory, F., Guipert, G., Mayda, S., 2014. Dating the *Homo erectus* bearing travertine from Kocabas (Denizli, Turkey) at at least 1.1 Ma. *Earth Planet. Sci. Lett.* 390, 8–18.
- Levin, N.E., 2015. Environment and climate of early human evolution. *Annu. Rev. Earth Planet. Sci.* 43, 405–429.  
<https://doi.org/10.1146/annurev-earth-060614-105310>
- Liddy, H.M., Feakins, S.J., Tierney, J.E., 2016. Cooling and drying in northeast Africa across the Pliocene. *Earth Planet. Sci. Lett.* 449, 430–438.  
<https://doi.org/10.1016/j.epsl.2016.05.005>
- Loth, P.E., Prins, H.H.T., 1986. Spatial patterns of the landscape and vegetation of Lake Manyara National Park. *ITC J.* 2, 115–130.
- Louchart, A., Wesselman, H., Blumenshine, R.J., Hlusko, L.J., Njau, J.K., Black, M.T., Asnake, M., White, T.D., 2009. Taphonomic, avian, and small-vertebrate indicators of *Ardipithecus ramidus* habitat. *SCIENCE* 326.  
<https://doi.org/10.1126/science.1175823>
- Magill, C.R., Ashley, G.M., Domínguez-Rodrigo, M., Freeman, K.H., 2016.

- Dietary options and behavior suggested by plant biomarker evidence in an early human habitat. *Proc. Natl. Acad. Sci.* 113, 2874–2879.
- McCool, J.-P., 2018. Carbonates as evidence for groundwater discharge to the Nile River during the Late Pleistocene and Holocene. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2018.09.026>
- McDougall, I., Brown, F.H., Vasconcelos, P.M., Cohen, B.E., Thiede, D.S., Buchanan, M.J., 2012. New single crystal 40Ar/39Ar ages improve time scale for deposition of the Omo Group, Omo-Turkana Basin, East Africa. *J. Geol. Soc.* 169, 213–226.
- McHenry, H.M., Brown, C.C., McHenry, L.J., 2007. Fossil hominin ulnae and the forelimb of *Paranthropus*. *Am. J. Phys. Anthropol. Off. Publ. Am. Assoc. Phys. Anthropol.* 134, 209–218.
- McKee, J.K., 1994. Catalogue of fossil sites at the Buxton Limeworks, Taung. *Palaeontol. Afr.* 31, 73–81.
- McKee, J.K., Kuykendall, K.L., 2016. The Dart Deposits of the Buxton Limeworks, Taung, South Africa, and the context of the Taung *Australopithecus* fossil. *J. Vertebr. Paleontol.* 36, e1054937.
- Moussa, A., Novello, A., Lebatard, A.-E., Decarreau, A., Fontaine, C., Barboni, D., Sylvestre, F., Bourlès, D.L., Paillès, C., Buchet, G., others, 2016. Lake Chad sedimentation and environments during the late Miocene and Pliocene: New evidence from mineralogy and chemistry of the Bol core sediments. *J. Afr. Earth Sci.* 118, 192–204. <https://doi.org/10.1016/j.jafrearsci.2016.02.023>
- Murphy, N.P., Guzik, M.T., Cooper, S.J.B., Austin, A.D., 2015. Desert spring refugia: museums of diversity or evolutionary cradles? *Zool. Scr.* 44, 693–701. <https://doi.org/10.1111/zsc.12129>
- Neumann, K., Fahmy, A.G., Müller-Scheebel, N., Schmidt, M., 2017. Taxonomic, ecological and palaeoecological significance of leaf phytoliths in West African grasses. *Quat. Int.* 434, 15–32. <https://doi.org/10.1016/j.quaint.2015.11.039>
- Nicoll, K., Giegengack, R., Kleindienst, M.R., 1999. Petrogenesis of artifact-bearing fossil-spring tufa deposits from Kharga Oasis, Egypt. *Geoarchaeology-Int. J.* 14, 849–863.
- Novello, A., 2012. Les phytolithes, marqueurs des environnements mio-pliocènes du Tchad. reconstitution à partir du signal environnemental des phytolithes dans l’Afrique subsaharienne actuelle (Doctorat). Université de Poitiers, Poitiers, France.
- Novello, A., Barboni, D., Berti-Equille, L., Mazur, J.-C., Poilecot, P., Vignaud, P., 2012. Phytolith signal of aquatic plants and soils in Chad, Central Africa. *Rev. Palaeobot. Palynol.* 178, 43–58. <https://doi.org/10.1016/j.revpalbo.2012.03.010>
- Novello, A., Barboni, D., Sylvestre, F., Lebatard, A.-E., Paillès, C., Bourlès, D.L., Likius, A., Mackaye, H.T., Vignaud, P., Brunet, M., 2017. Phytoliths indicate significant arboreal cover at Sahelanthropus type locality TM266 in northern Chad and a decrease in later sites. *J. Hum. Evol.* 106, 66–83. <https://doi.org/10.1016/j.jhevol.2017.01.009>
- Novello, A., Lebatard, A.-E., Moussa, A., Barboni, D., Sylvestre, F., Bourlès, D.L., Paillès, C., Buchet, G., Decarreau, A., Düringer, P., Ghienne, J.-F., Maley, J., Mazur, J.-C., Roquin, C., Schuster, M., Vignaud, P., 2015. Diatom, phytolith, and pollen records from a  $^{10}\text{Be}/^{9}\text{Be}$  dated lacustrine succession in the Chad basin: Insight on the Miocene–Pliocene paleoenvironmental changes in Central Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 430, 85–103.
- Olago, D., Opere, A., Barongo, J., 2009. Holocene palaeohydrology, groundwater and climate change in the lake basins of

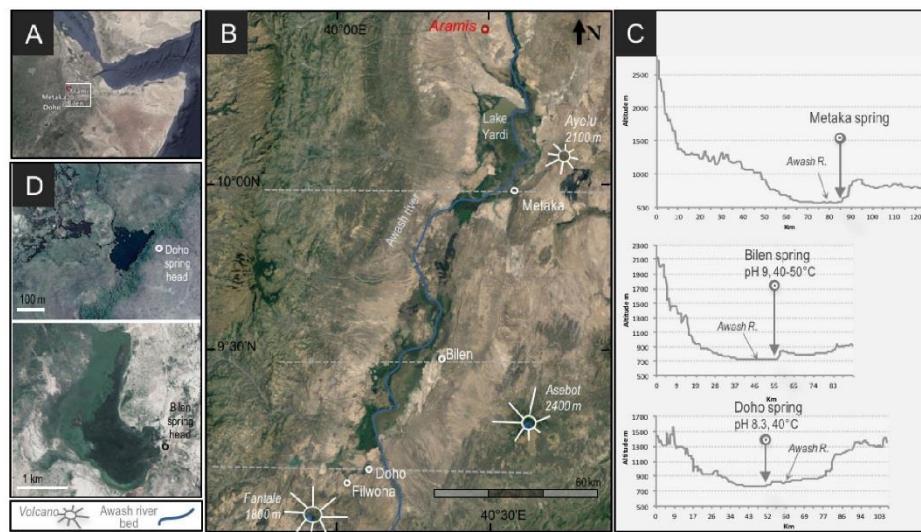
- the Central Kenya Rift. *Hydrol. Sci. J.* 54, 765–780.  
<https://doi.org/10.1623/hysj.54.4.765>
- O'Regan, H.J., Wilkinson, D.M., Marston, C.G., 2016. Hominin home ranges and habitat variability: Exploring modern African analogues using remote sensing. *J. Archaeol. Sci. Rep.* 9, 238–248. <https://doi.org/10.1016/j.jasrep.2016.06.043>
- Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., Simons, A., 2009. Agroforestry database: a tree species reference and selection guide version 4.0. World Agrofor. Cent. ICRAF Nairobi KE.
- Owen, R.B., Renaut, R.W., Hover, V.C., Ashley, G., Muasya, A.M., 2004. Swamps, springs and diatoms: wetlands of the semi-arid Bogoria-Baringo Rift, Kenya. *Hydrobiologia* 518, 59–78. <https://doi.org/10.1023/B:HYDR.0000025057.62967.2c>
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371.
- Pickford, M., 1995. Fossil land snails of East Africa and their palaeoecological significance. *J. Afr. Earth Sci.* 20, 167–226.
- Pickford, M., Senut, B., 2001. The geological and faunal context of Late Miocene hominid remains from Lukeino, Kenya. *Comptes Rendus Académie Sci.-Ser. IIa-Earth Planet. Sci.* 332, 145–152.
- Pigati, J.S., Rech, J.A., Quade, J., Bright, J., 2014. Desert wetlands in the geologic record. *Earth-Sci. Rev.* 132, 67–81. <https://doi.org/10.1016/j.earscirev.2014.02.001>
- Porat, N., Chazan, M., Grün, R., Aubert, M., Eisenmann, V., Kolska Horwitz, L., 2010. New radiometric ages for the Fauresmith industry from Kathu Pan, southern Africa: Implications for the Earlier to Middle Stone Age transition. *J. Archaeol. Sci.* 37, 269–283.
- Potts, R., 2013. Hominin evolution in settings of strong environmental variability. *Quat. Sci. Rev.* 73, 1–13. <https://doi.org/10.1016/j.quascirev.2013.04.003>
- Pruettz, J.D., Bertolani, P., 2009. Chimpanzee (*Pan troglodytes verus*) behavioral responses to stresses associated with living in a savanna-mosaic environment: Implications for hominin adaptations to open habitats. *PaleoAnthropology* 252–262. <https://doi.org/10.4207/PA.2009.ART33>
- Quade, J., Cerling, T.E., Bowman, J.R., 1989. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. *Nature* 342, 163–166. <https://doi.org/10.1038/342163a0>
- R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reynolds, S.C., Bailey, G.N., King, G.C.P., 2011. Landscapes and their relation to hominin habitats: Case studies from Australopithecus sites in eastern and southern Africa. *J. Hum. Evol.* 60, 281–298. <https://doi.org/10.1016/j.jhevol.2010.10.001>
- Reynolds, S.C., Marston, C.G., Hassani, H., King, G.C., Bennett, M.R., 2016. Environmental hydro-refugia demonstrated by vegetation vigour in the Okavango Delta, Botswana. *Sci. Rep.* 6, 35951.
- Reynolds, S.C., Wilkinson, D.M., Marston, C.G., O'Regan, H.J., 2015. The ‘mosaic habitat’ concept in human evolution: past and present. *Trans. R. Soc. South Afr.* 70, 57–69.
- Richerson, P.J., Bettinger, R.L., Boyd, R., 2008. Evolution on a Restless Planet: Were Environmental Variability and Environmental Change Major Drivers of Human Evolution?, in: *Handbook of Evolution*. Wiley-VCH Verlag GmbH, pp. 223–242.
- Roche, D., Ségalen, L., Senut, B., Pickford,

- M., 2013. Stable isotope analyses of tooth enamel carbonate of large herbivores from the Tugen Hills deposits: Palaeoenvironmental context of the earliest Kenyan hominids. *Earth Planet. Sci. Lett.* 381, 39–51.
- Rooney, T.O., 2017. The Cenozoic magmatism of East-Africa: Part I — Flood basalts and pulsed magmatism. *Lithos* 286–287, 264–301. <https://doi.org/10.1016/j.lithos.2017.05.014>
- Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J., Hill, D.J., 2008. A new global biome reconstruction and data-model comparison for the Middle Pliocene. *Glob. Ecol. Biogeogr.* 17, 432–447. <https://doi.org/10.1111/j.1466-8238.2008.00381.x>
- Schuster, M., 2006. The Age of the Sahara Desert. *Science* 311, 821–821. <https://doi.org/10.1126/science.1120161>
- Semaw, S., Simpson, S.W., Quade, J., Renne, P.R., Butler, R.F., McIntosh, W.C., Levin, N.E., Dominguez-Rodrigo, M., Rogers, M.J., 2005. Early Pliocene hominids from Gona, Ethiopia. *Nature* 301–305.
- Senut, B., 2006. Bipédie et climat. *Comptes Rendus Palevol* 5, 89–98.
- Shipman, P., Harris, J.M., 1988. Habitat preference and paleoecology of *Australopithecus boisei* in Eastern Africa. *Evol. Hist. “robust” Australopithecines* 343–381.
- Smith, J.R., Giegengack, R., Schwarcz, H.P., McDonald, M.M.A., Kleindienst, M.R., Hawkins, A.L., Churcher, C.S., 2004. A reconstruction of Quaternary pluvial environments and human occupations using stratigraphy and geochronology of fossil-spring tufas, Kharga Oasis, Egypt. *Geoarchaeology* 19, 407–439. <https://doi.org/10.1002/gea.20004>
- Smith, J.R., Hawkins, A.L., Asmerom, Y., Polyak, V., Giegengack, R., 2007. New age constraints on the Middle Stone Age occupations of Kharga Oasis, Western Desert, Egypt. *J. Hum. Evol.* 52, 690–701.
- Sockol, M.D., Raichlen, D.A., Pontzer, H., 2007. Chimpanzee locomotor energetics and the origin of human bipedalism. *Proc. Natl. Acad. Sci.* 104, 12265–12269.
- Stauffer, F.W., Ouattara, D., Stork, A.L., 2014. Palmae, in: Lebrun, J.-P., Stork, A.L. (Eds.), *Tropical African Flowering Plants: Monocotyledons 2. Conservatoire et Jardin botaniques de la Ville de Genève*, Switzerland, pp. 326–354.
- Staver, A.C., Archibald, S., Levin, S.A., 2011. The global extent and determinants of savanna and forest as alternative biome states. *Science* 334, 230–232. <https://doi.org/10.1126/science.1210465>
- Steudel-Numbers, K.L., Tilkens, M.J., 2004. The effect of lower limb length on the energetic cost of locomotion: implications for fossil hominins. *J. Hum. Evol.* 47, 95–109.
- Stevens, L.E., Meretsky, V.J., 2008. Aridland springs in North America: ecology and conservation. University of Arizona Press.
- Suhling, F., Sahlén, G., Martens, A., Marais, E., Schütte, C., 2006. Dragonfly Assemblages in Arid Tropical Environments: A Case Study from Western Namibia. *Biodivers. Conserv.* 15, 311–332. <https://doi.org/10.1007/s10531-005-2007-6>
- Suwa, G., Ambrose, S.H., 2014. Reply to Cerling et al. *Curr. Anthropol.* 55, 473–474.
- Suwa, G., Kono, R.T., Simpson, S.W., Asfaw, B., Lovejoy, C.O., White, T.D., 2009. Paleobiological Implications of the *Ardipithecus ramidus* Dentition. *Science* 326, 94–99. <https://doi.org/10.1126/science.1175824>
- Swedell, L., 2002. Ranging behavior, group size and behavioral flexibility in Ethiopian hamadryas baboons (*Papio hamadryas hamadryas*). *Folia Primatol.* (Basel) 73, 95–103.
- Szaro, R.C., Jakle, M.D., 1985. Avian use of a desert riparian island and its adjacent

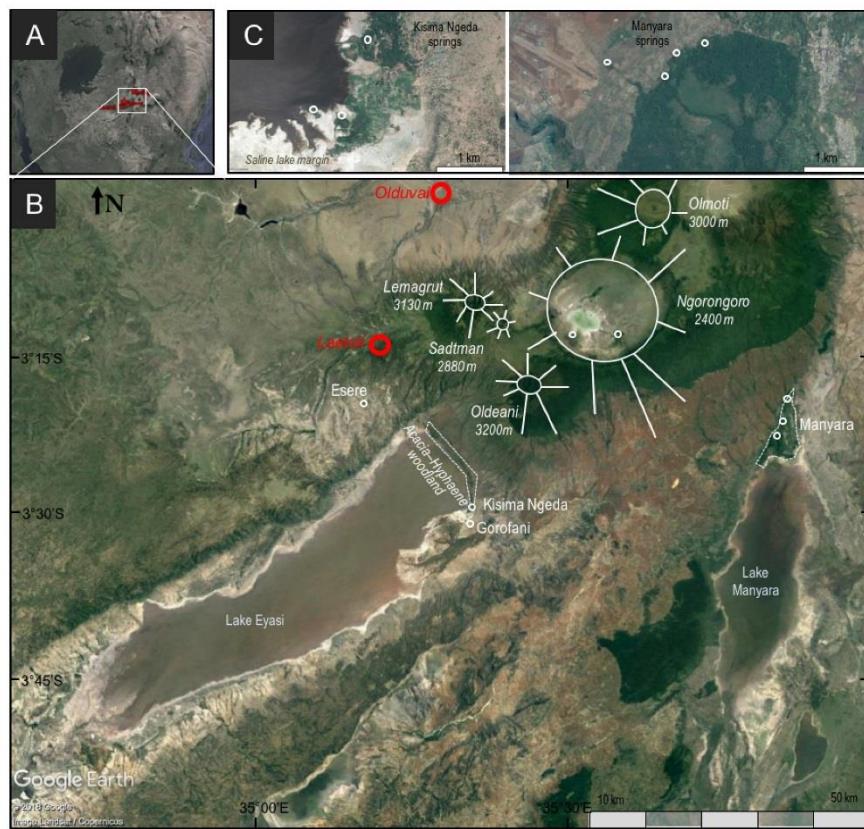
- scrub habitat. *The Condor* 87, 511–519. <https://doi.org/10.2307/1367948>
- Tryon, C.A., Faith, J.T., Peppe, D.J., Keegan, W.F., Keegan, K.N., Jenkins, K.H., Nightingale, S., Patterson, D., van Plantinga, A., Driese, S., Johnson, C.R., Beverly, E.J., 2014. Sites on the landscape: Paleoenvironmental context of late Pleistocene archaeological sites from the Lake Victoria basin, equatorial East Africa. *Quat. Int.* 331, 20–30.
- Tryon, C.A., Peppe, D.J., Faith, J.T., Van Plantinga, A., Nightingale, S., Ogando, J., Fox, D.L., 2012. Late Pleistocene artefacts and fauna from Rusinga and Mfangano islands, Lake Victoria, Kenya. *Azania Archaeol. Res. Afr.* 47, 14–38. <https://doi.org/10.1080/0067270X.2011.647946>
- Uno, K.T., Polissar, P.J., Jackson, K.E., deMenocal, P.B., 2016. Neogene biomarker record of vegetation change in eastern Africa. *Proc. Natl. Acad. Sci.* 113, 6355–6363. <https://doi.org/10.1073/pnas.1521267113>
- Van Plantinga, A., 2011. Geology of the Late Pleistocene artifact-bearing Wasiriyah beds at the Nyamita Locality, Rusinga Island, Kenya (Master's Thesis). Baylor University, Waco, TX.
- Viallet, A., Guipert, G., Alcicek, M.C., 2012. Homo erectus found still further west: reconstruction of the Kocabas cranium (Denizli, Turkey). *Comptes Rendus Palevol* 11, 89–95.
- Vignaud, P., Duringer, P., Mackaye, H.T., Likius, A., Blondel, C., Boisserie, J.-R., de Bonis, L., Eisenmann, V., Etienne, M.-E., Geraads, D., Guy, F., Lehmann, T., Lihoreau, F., Lopez-Martinez, N., Mourer-Chauviré, C., Otero, O., Rage, J.-C., Schuster, M., Viriot, L., Zazzo, A., Brunet, M., 2002. Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* 418, 152–155. <https://doi.org/10.1038/nature00880>
- Viljoen, P.J., Bothma, J., Du, P., 1990. Daily movements of desert-dwelling elephants in the northern Namib Desert. *South Afr. J. Wildl. Res.* 20, 69–72.
- Vrba, E.S., 1980. The significance of bovid remains as indicators of environment and predation patterns, in: Behrensmeyer, A.K., Hill, A.P. (Eds.), *Fossils in the Making*. University of Chicago Press, Chicago, pp. 247–272.
- Wendorf, F., Schild, R., Close, A., Associates (Eds.), 1993. Egypt during the last interglacial: The Middle Paleolithic of Bir Tarfawi and Bir Sahara East. New York: Plenum. Plenum Press, New York.
- White, T.D., Ambrose, S.H., Suwa, G., Su, D.F., DeGusta, D., Bernor, R.L., Boisserie, J.-R., Brunet, M., Delson, E., Frost, S., Garcia, N., Giaourtsakis, I.X., Haile-Selassie, Y., Howell, F.C., Lehmann, T., Likius, A., Pehlevan, C., Saegusa, H., Semprebon, G., Teaford, M., Vrba, E., 2009a. Macrovertebrate Paleontology and the Pliocene Habitat of *Ardipithecus ramidus*. *SCIENCE* 326, 87–93. <https://doi.org/10.1126/science.1175822>
- White, T.D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C.O., Suwa, G., WoldeGabriel, G., 2009b. *Ardipithecus ramidus* and the paleobiology of early hominids. *Science* 326, 75–86. <https://doi.org/10.1126/science.1175802>
- White, T.D., Lovejoy, C.O., Asfaw, B., Carlson, J.P., Suwa, G., 2015a. Neither chimpanzee nor human, *Ardipithecus* reveals the surprising ancestry of both. *Proc. Natl. Acad. Sci.* 201403659.
- White, T.D., Suwa, G., Ambrose, S.H., 2015b. Reply to Cerling et al. *Curr. Anthropol.* 56, 447–448.
- Williams, M.A.J., Bishop, P.M., Dakin, F.M., Gillespie, R., 1977. Late Quaternary lake levels in southern Afar and the adjacent Ethiopian Rift. *Nature* 267, 690–693.
- WoldeGabriel, G., Ambrose, S.H., Barboni, D., Bonnefille, R., Bremond, L., Currie, B., DeGusta, D., Hart, W.K., Murray, A.M., Renne, P.R., Jolly-Saad, M.C., Stewart, K.M., White, T.D., 2009. *M.C.*, Stewart, K.M., White, T.D., 2009.

The geological, isotopic, botanical, invertebrate, and lower vertebrate surroundings of *Ardipithecus ramidus*. Science 326. <https://doi.org/10.1126/science.1175817>  
Zhang, Z., Ramstein, G., Schuster, M., Li,

C., Contoux, C., Yan, Q., 2014. Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. Nature 513, 401.



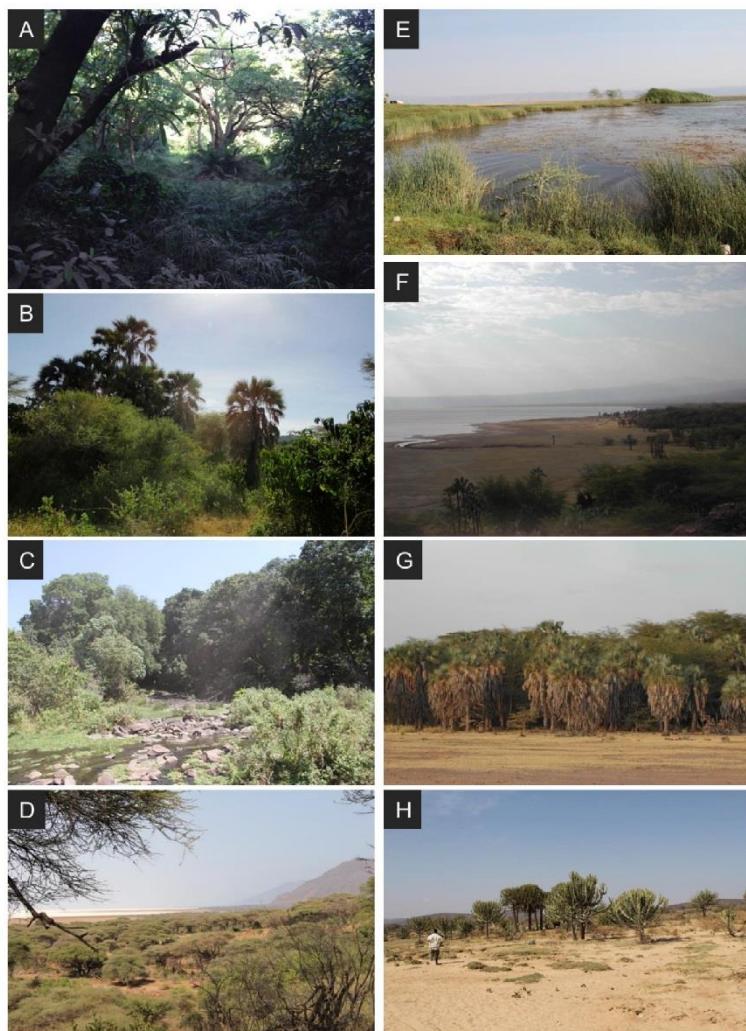
**Figure 1.** Springs in the Awash Valley, Afar region, Ethiopia. (A) Position of the Awash Valley. (B) Satellite view (Landsat / Copernicus image from GoogleEarth, 2018) with position of volcanoes, Awash River, and the four spring sites visited in 2018: Filwoha, Doho, Bilen and Metaka. (C) Elevation profiles from west to east cutting through the rift valley at the latitude of Metaka, Bilen and Doho. Elevation profiles were obtained using Zonums free software. (D) Close-ups show sharp vegetation changes at the spring sites.



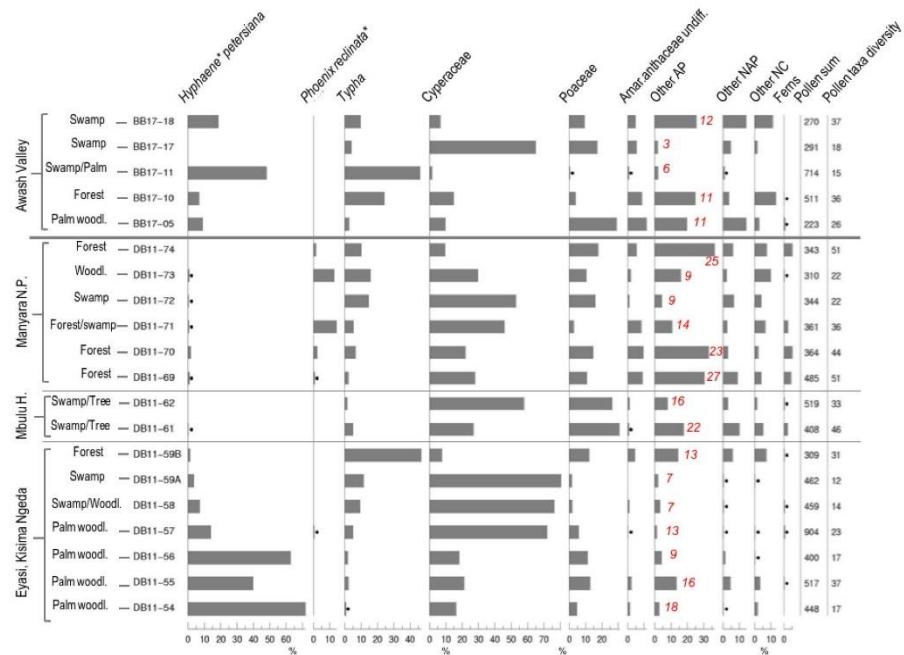
**Figure 2.** Springs in the Ngorongoro – Lake Eyasi – Lake Manyara region, north Tanzania. (A) General view. (B) Satellite view (Landsat / Copernicus image from GoogleEarth, 2018) with position of volcanoes, saline lakes Eyasi and Manyara, and the spring sites visited several times since 2012, mainly Kisima Ngeda and Manyara spring forest. Red circles indicate paleontological sites of Olduvai and Laetoli. (C) Close-ups show sharp vegetation changes at the spring sites, located at the foothills of Oldeani and Ngorongoro Highlands.



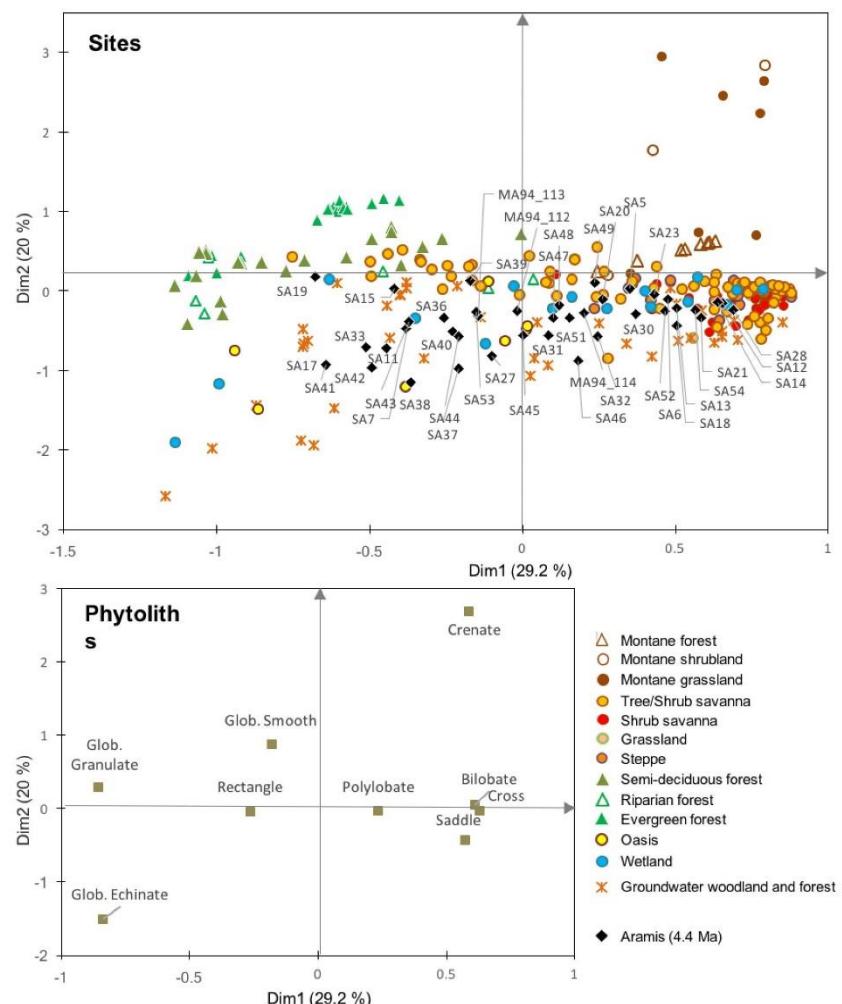
**Figure 3.** Photographs of modern spring palm woodlands of the Awash Valley, Ethiopia. A) Palm spring woodland with abundant grasses and some *Acacia* and Capparidaceae shrubs, Filwoha area, Awash National Park; B-D) Doho spring area, B) *Hyphaene thebaica* palms, *Acacia nilotica* and exotic *Prosopis* trees in the background at the base of the basaltic high, Cyperaceae on damp to wet soils; C) view from the basaltic high showing the extent of the spring woodland, patches of *Typha* and of grasses occur within the palm woodland, D) *Hyphaene thebaica* palm trees growing at the spring head, Doho Lodge.



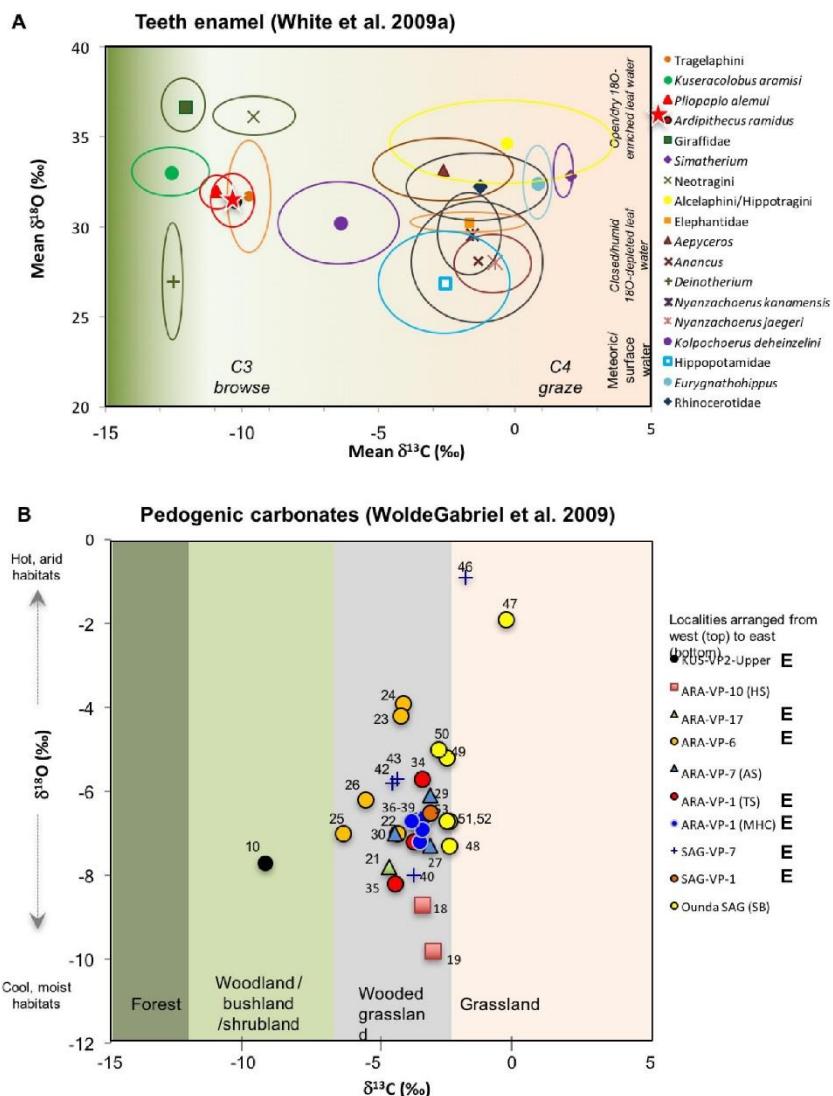
**Figure 4.** Photographs of the modern azonal (spring) and zonal vegetation in the northwestern edge of Lake Manyara (A-D) and of northeastern edge of Lake Eyasi (E-H), Tanzania. A: Spring forest with abundant evergreen trees, understory includes sedges (foreground) and *Phoenix reclinata* palm treelets (background)– ground was damp; B: Woodland with *Hyphaene petersiana* palm trees and various shrubs at the southern edge of the spring forest –ground dry; C: Evergreen forest growing along one of the spring-fed rivers reaching Lake Manyara floodplain; D: *Acacia* – *Commiphora* thicket-woodland on the distal floodplain of Lake Manyara; E: Freshwater wetland on Lake Eyasi floodplain near Kisima Ngenda spring with *Typha*, sedges, and the small halophytic tree *Sesbania sesban*; F: Lake Eyasi floodplain with *Acacia xanthophloea* and *Hyphaene petersiana* woodland on the right; G: close-up on the *Acacia xanthophloea* and *Hyphaene petersiana* woodland; H: zonal scrubland with rare grass patches and the cactoid *Euphorbia candelabrum*. Photos by the authors, except F and G (M. Dominguez-Rodrigo).



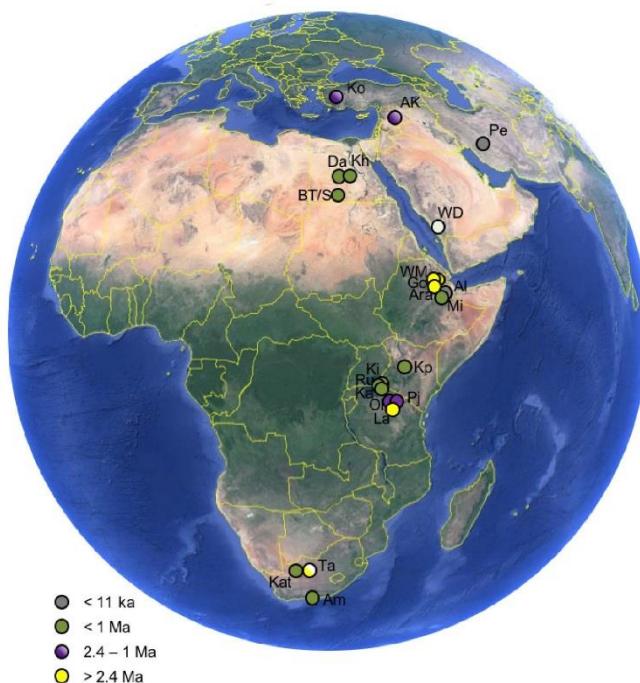
**Figure 5.** Pollen spectra from modern spring sites in Ethiopia and Tanzania. Numbers in red indicate pollen diversity among arboreal taxa. \* means type, *Hyphaene*-type *petersiana* for example. Percentages <0.5% are marked by a black dot. Details regarding site location and pollen counts are given in SOM1.



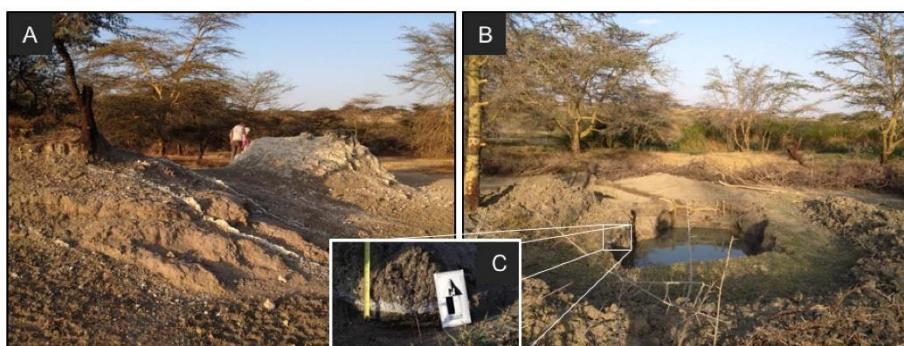
**Figure 6.** Correspondence analysis ordination diagram of 265 modern sites and 9 phytolith morphotypes. Fossil Aramis data points used as supplementary (passive) variables in the CA are shown as black diamonds.



**Figure 7.** Aramis isotopic data  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  on teeth enamel and pedogenic carbonates. Data and inferred C<sub>3</sub>/C<sub>4</sub> diets are from White et al. (2009a) and WoldeGabriel et al. (2009). Probable vegetation type inferred from  $\delta^{13}\text{C}$  values according to Cerling et al. (2011).



**Figure 8.** Position of the sites with early hominin remains and/or artifacts found associated with coeval geological evidence for springs. Google Earth map (Data SIO, NOAA, U.S. Navy, NGA, GEBCO; Image IBCAO, Image Landsat / Copernicus).



**Figure 9.** Photographs of fossil and modern spring at Esere, near Laetoli area, north Tanzania. (A) Carbonate mount with carbonate-encrusted plant remains, which attest to the presence of a fossil spring (dating of carbonate, in progress). (B) Modern, active spring adjacent to the carbonate mount. In the foreground, the water hole dug by herders shows that groundwater is <1 m below the surface. In the background, where spring water naturally reaches the surface, *Typha* reeds and sedges occur. (C) Micritic carbonate bed within the soil.

## **Annexe paper 1.2 - Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia**



### **Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia**

Beverly Saylor, Luis Gibert, Alan Deino, Mulugeta Alene, Naomi Levin,  
Stephanie Melillo, Mark Peaple, Sarah Feakins, Benjamin Bourel, Doris  
Barboni, et al.

#### ► To cite this version:

Beverly Saylor, Luis Gibert, Alan Deino, Mulugeta Alene, Naomi Levin, et al.. Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia. *Nature*, Nature Publishing Group, 2019, 573 (7773), pp.220-224. 10.1038/s41586-019-1514-7 . hal-02286089

**HAL Id: hal-02286089**

<https://hal.archives-ouvertes.fr/hal-02286089>

Submitted on 21 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[This is the subedited preproof of the paper, not including figures. It was corrected in line with the instructions and comments of the proofreader]

The full paper with figures can be seen read following this Springer Nature SharedIt link:  
<https://rdcu.be/bU0kL>

## **Age and context of mid-Pliocene hominin cranium from Woranso-Mille, Ethiopia**

Beverly Z. Saylor<sup>1</sup>, Luis Gibert<sup>2</sup>, Alan Deino<sup>3</sup>, Mulugeta Alene<sup>4</sup>, Naomi E. Levin<sup>5</sup>, Stephanie M. Melillo<sup>6</sup>, Mark D. Peaple<sup>7</sup>, Sarah J. Feakins<sup>7</sup>, Benjamin Bourel<sup>8</sup>, Doris Barboni<sup>8</sup>, Alice Novello<sup>8</sup>, Florence Sylvestre<sup>8</sup>, Stanley A. Mertzman<sup>9</sup> & Yohannes Haile-Selassie<sup>10\*</sup>

<sup>1</sup>Department of Earth, Environmental and Planetary Sciences, Case Western Reserve University, Cleveland, OH, USA.

<sup>2</sup>Departament de Mineralogia, Petrologia i Geologia Aplicada Facultat de Ciències de la Terra, Universitat de Barcelona, Barcelona, Spain.

<sup>3</sup>Berkeley Geochronology Center, Berkeley, CA, USA.

<sup>4</sup>School of Earth Sciences, Addis Ababa University, Addis Ababa, Ethiopia.

<sup>5</sup>Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA.

<sup>6</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

<sup>7</sup>Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA.

<sup>8</sup>Aix-Marseille University, CNRS, IRD, INRA, Collège de France, CEREGE, Aix-en-Provence, France.

<sup>9</sup>Department of Earth and Environment, Franklin and Marshall College, Lancaster, PA, USA.

<sup>10</sup>Cleveland Museum of Natural History, Cleveland, OH, USA.

\*email: [yhaileselassie@cmnh.org](mailto:yhaileselassie@cmnh.org)

**A fossil hominin cranium was discovered in mid-Pliocene deltaic strata in the Godaya Valley of the northwestern Woranso-Mille study area in Ethiopia. Here we show that analyses of chemically correlated volcanic layers and the palaeomagnetic stratigraphy, combined with Bayesian modelling of dated tuffs, yield an age range of  $3.804 \pm 0.013$  to  $3.777 \pm 0.014$  million years old ( $1\sigma$ ) for the deltaic strata and the fossils that they contain.**

**We also document deposits of a perennial lake beneath the deltaic sequence. Mammalian fossils associated with the cranium represent taxa that were widespread at the time, and data from botanical remains indicate that the vegetation in the lake and delta catchment was predominantly dry shrubland with varying proportions of grassland, wetland and riparian forest. In addition, we report high rates of sediment accumulation and depositional features that are typical of a steep topographic relief and differ from younger Woranso-Mille fossil localities reflecting the influence of active rift processes on the paleolandscape.**

Until recently, field work in the Woranso-Mille area (Afar, Ethiopia) has concentrated on exposures along the Mille River that include more than 120 m of fossiliferous sedimentary and volcanic strata that are between 3.8 and 3.2 million years (Myr) old. East of Korsi Dora (Fig. 1a), sedimentary sequences that are younger than the extra-regional approximately 3.57-Myr-old Kilaytoli tuff<sup>1</sup> have yielded specimens of at least two hominin species, *Australopithecus afarensis* and *Australopithecus deyiremeda*<sup>2–5</sup>. West of Korsi Dora, near the Waki-Mille confluence, volcanic and sedimentary layers between the approximately 3.76-Myr-old<sup>1</sup> Mille tuff sequence and the Kilaytoli tuff yielded dentognathic fossils that seem to be transitional between *A. anamensis* and *A. afarensis*<sup>6</sup>. A fossil cranium (MRD-VP-1/1), discovered in 2016 and reported in a companion paper<sup>7</sup>, was recovered 6 km northwest of the Waki-Mille confluence in the Godaya Valley, where the section extends 55 m below the Mille tuff sequence and 60–125 m above it (Fig. 1b). The section below the Mille tuff sequence (Fig. 2) overlaps in age with the lowermost Hadar Formation<sup>8–10</sup> and the uppermost Sagantole Formation<sup>11–13</sup> of the neighbouring Awash Valley (Fig. 1a). Even though these formations are known for their Pliocene hominin fossils, they have yielded only a single hominin specimen between 4.1 and 3.6 Myr old<sup>14,15</sup>. Outside the Afar, the hominin record from this time interval is also limited<sup>16–20</sup>.

### **Stratigraphic setting and age of the fossils**

In the study area, the Godaya River passes between the high Guda plateau and the lower Miro Dora and Am-Ado plateaus (Fig. 1c). The low plateaus are covered by basalt flows that have been geochemically identified as the Am-Ado basalt group, one of six basalt groups in the Woranso-Mille area<sup>21</sup> (Supplementary Table 1). As these basalt groups are stratigraphically compact and chemically distinguishable, they form reliable chronostratigraphic markers (Fig. 1).

The MRD-VP-1 fossil locality is at the bottom of the Miro Dora Plateau, around 23 m stratigraphically beneath the Am-Ado basalt group (Extended Data Fig. 1a). The collection horizon consists of low outcrops of coarse sandstone, part of a succession of deltaic siltstone and sandstone (section S4 in Fig. 2, see ‘Lake and delta sequence’). Most fossil specimens, including MRD-VP-1/1, show no sign of abrasion, although a few uncollected fragments have rolled edges, which is indicative of transport. The coarse sand matrix in the region of the orbits indicates MRD-VP-1/1 was buried in the coarse sandstone of the MRD-VP-1 locality (Extended Data Fig. 1b), although possibly after burial elsewhere had left a hard, fine-sand matrix in the nasal opening, cranial base and parts of the hard palate.

Tuffs above the Am-Ado basalt group are younger than the extra-regional approximately 3.66-Myr-old Waki tuff<sup>22</sup> based on stratigraphic and geochemical correlations to tuffs near the Waki-Mille confluence<sup>22</sup> (Extended Data Fig. 2 and Supplementary Table 2). These tuffs are preserved (1) northwest of MRD-VP-1, at and near the measured section S6; (2) southeast of MRD-VP-1 at Am-Ado; and (3) east of MRD-VP-1, along section S1, below the extra-regional approximately 3.4-Myr-old<sup>23</sup> Sidi Hakoma tuff (Fig. 1).

Between the Am-Ado basalt group and the fossiliferous deltaic sequence below is a tuffaceous interval (Fig. 1). This interval is thickest (around 22 m) at S2 (Lehaysule Gera hill; Extended Data Fig. 1c) where both the Am-Ado basalt flows and the sandstone of the deltaic sequence have pinched out. There it consists of metre-scale felsic pumice tuffs with interbedded basaltic pumice tuffs, scoria, pumice siltstone and siltstone, and includes the Mille tuff sequence, which is recognized on the basis of physical characteristics and stratigraphic position (Fig. 2). Tuffs below the Mille tuff sequence are typically devitrified; however, at S2 and S5 (3 km northwest of S2; Extended Data Fig. 1d) tuffs within and above the Mille tuff sequence yielded glass shards suitable for geochemical analysis (Supplementary Table 2). These tuffs are compositionally and stratigraphically similar to the Mille tuff sequence, ‘bimodal tuff’ (basaltic and rhyolitic) and Araskimiro tuff, which are closely associated tuffs near the Waki-Mille confluence that underlie the approximately 3.66-Myr-old Waki tuff and have statistically indistinguishable  $^{40}\text{Ar}/^{39}\text{Ar}$  ages<sup>1,22</sup> of 3.77–3.76 Myr (Fig. 1b and Extended Data Fig. 2). These geochemical correlations confirm the recognition of the Mille tuff sequence in the Godaya Valley, including across a northwestward facies change from fallout pumice tuffs in sections S2

and S3 to pumice siltstone in sections S4, S5 and S6 (Fig. 2). The age of the Mille tuff provides a minimum constraint<sup>1</sup> of around 3.76 Myr for the fossil horizon.

At the bottom of the tuffaceous interval is a coarse, locally thick (0–3 m) pumice lapillistone fallout tuff that is informally referred to as T9 (Fig. 2). T9 is underlain by deltaic sediments, below which there is lacustrine claystone with thin (mostly less than 10 cm) intercalated tuffs. These lakebed tuffs are best exposed at S2 and S6 where the lithologic characteristics, spacing and uniformity of thickness permit bed-for-bed correlations (tuff units T1–T8) across more than 4 km (Extended Data Fig. 3). Thus, although the T1–T8 tuffs are not exposed at S4, they can confidently be projected below the fossil horizon there (Figs. 1d, 2). Two tuffs from S2 were dated using the  $^{40}\text{Ar}/^{39}\text{Ar}$  laser incremental-heating dating method, providing bracketing constraints on the age of the delta sequence. These are WM12/LHG-6, a sample of the T9 tuff, and WM18/LHG-1, a sample of a 2-cm thick tuff, 30 cm below T1 (Fig. 2, section S2).

The  $^{40}\text{Ar}/^{39}\text{Ar}$  dating experiments involved single-crystal laser incremental heating of individual phenocrysts. In total, 10 out of 13 incremental-heating experiments from WM12/LHG-6 and all 8 experiments from WM18/LHG-1 yielded apparent-age plateaus (Extended Data Fig. 4 and Supplementary Table 3). The argon isotope data from the plateau steps were plotted on inverse-isochron  $^{36}\text{Ar}/^{40}\text{Ar}$  and  $^{39}\text{Ar}/^{40}\text{Ar}$  correlation diagrams (Extended Data Fig. 5), the results of which are displayed as age–probability density spectra in Extended Data Fig. 6, combining all individual crystal isochron results for each sample into a single population. The weighted-mean age of the quasi-symmetrical, quasi-Gaussian modes representing the two populations is  $3.816 \pm 0.021$  Myr old for WM12/LHG-6 ( $1\sigma$  including the error in  $J$ , the neutron-fluence parameter; mean square weighted deviation (MSWD) = 1.46) and  $3.786 \pm 0.033$  Myr old for WM18/LHG-1 (MSWD = 0.07). Although apparently inverted (the lower sample age is younger than the higher), the ages are indistinguishable given the large uncertainties owing to low K abundance in plagioclase for WM12/LHG-6 and small grain size (around 200–400  $\mu\text{m}$ ) for WM18/LHG-1. These results are consistent with an age<sup>1</sup> of around 3.76 Myr for the Mille tuff sequence (Fig. 2).

We computed a Bayesian stratigraphic age model for section S2 that extends from the lowermost dated tuff (WM18/LHG-1) to the Mille tuff sequence, as well as a similar model for

the laterally equivalent but thicker interval at S6 (Extended Data Fig. 2). These models allow interpolation of ages for the bottom ( $3.804 \pm 0.013$  Myr at S2 and  $3.802 \pm 0.016$  Myr at S6) and top ( $3.798 \pm 0.014$  Myr at S2 and  $3.777 \pm 0.014$  Myr at S6) of the deltaic sediments, which yielded indistinguishable estimates and an age range of  $3.804 \pm 0.013$  Myr to  $3.777 \pm 0.014$  Myr for the deltaic sediments and the cranium.

Palaeomagnetic reversal stratigraphy supports this age assignment. A 100-m reverse–normal–reverse palaeomagnetic sequence at S6 includes a more than 64 m-thick reverse zone at the bottom that extends from 46 m below to 18 m above the Am-Ado basalt group, encompassing the T1–T9 tuffs and the Mille tuff sequence (Fig. 1, Extended Data Fig. 7 and Supplementary Table 4). This zone is correlated with palaeomagnetic subchron C2Ar (which is 4.187–3.596 Myr old)<sup>24</sup>. Its lower boundary is not locally exposed, but its upper boundary with the normal C2An.3n (3.596–3.33 Myr old)<sup>24</sup> is identified 5 m above the Kilaytoli tuff, consistent with the position of this subchron boundary at Korsi Dora<sup>1,25</sup>. Reverse palaeomagnetic strata at S2 and S4 are consistent with the correlation of these sections with C2Ar (Fig. 2 and Extended Data Fig. 7). On the basis of the modelled ages of tuffs, the sediment accumulation rate for the interval between WM18/LHG-1 and the Mille tuff sequence is 320 m per Myr at S2 and 820 m per Myr at S6 (Supplementary Table 5). These high rates contrast with lower accumulation rates above the Mille tuff sequence at S6 (90–100 m per Myr for intervals up to the top of the C2An.3n normal zone) and S2 (170–210 m per Myr for intervals up to the Sidi Hakoma tuff).

### Lake and delta sequence

The sedimentary interval below the T9 tuff is 10.5 m (S2) to 51.5 m (S5) thick and consists of three facies: claystone, heterolithic interbeds and coarse to pebbly sandstone (Fig. 2). The claystone facies exceeds 32 m at S5, the lower part of which is greenish grey with carbonate concretions and beds of conglomerate, gastropods and sandstone. The conglomerate is matrix-supported and, along the drainage between S4 and S5, in beds up to 2 m thick with sand-filled root casts (Fig. 1). Clasts are sub-rounded, up to 15 cm across and diverse, including basalt, rhyolite, quartzite, sandstone and reworked carbonate nodules. The upper part of the claystone is greyish green and locally red, and contains the altered lakebed tuffs (WM18/LHG-1, T1–T8; Fig. 2). Claystone grades upward into 2–14 m of sandier upward, heterolithic interbeds. Decimetre-scale sandstone beds within this facies have sharp bases, normal grading and soft-sediment

deformation. Above the heterolithic facies is 0–13 m of coarse to pebbly sandstone; this is the facies that contained MRD-VP-1/1. At S3, S4 and S6, this sandstone is tabular and trough cross-bedded and planar stratified. Straight-crested foresets are 3–5.5 m high and, at MRD-VP-1 (S4), are traceable in satellite images over 600 m of outcrop, which reveal dominantly southeast to southwest palaeoflow directions (Fig. 1c and Extended Data Fig. 1e). Field measurements of palaeoflow directions range from northeast to southwest (Fig. 2).

Claystone is a product of slow settling in a lake environment, beyond the normal reach of coarser sediment, whereas the matrix-supported conglomerate is characteristic of debris flows, which can transport material for long distances even on shallow slopes or under water<sup>26</sup>. However, root casts in the conglomerate and an absence of clay in overlying gastropod sandstone indicate that these beds were associated with lake-margin settings that were subject to plant colonization and winnowing of fine sediment. By contrast, the lateral continuity of the lakebed tuffs (T1–T8) in the upper part of the claystone facies is consistent with suspension fallout on the floor of a perennial lake without evidence of winnowing, erosion or desiccation. The large, straight-crested foresets in the sandstone are typical of clinoforms that are formed by a Gilbert-type delta in which a river enters a lake and rapidly deposits its coarse sediment load<sup>27</sup>. Trough cross-bedding and planar stratification associated with, and above, the clinoforms are consistent with fluvial top sets (Fig. 2). Sedimentary structures in the heterolithic facies are typical of current- or density-generated flows, which occur in pro-delta and lake-margin settings, in front of and next to delta lobes<sup>27</sup>. The complex stratigraphy includes multi-leveled cross-sets that may reflect internal delta dynamics or repeated episodes of progradation, yet it is clear that before the T9 tuff, the delta had built southeastward above the T1–T8 tuffs to at least as far as MRD-VP-1 (S4), and that before the Mille tuff, a lobe had reached S3 (Extended Data Fig. 1f). From there, the deltaic sandstone thins northeastward towards S2, pinching out over approximately 1 km. Changes in lobe thickness and depositional settings along and across the delta axis can account for changes in the overlying tuffaceous interval. In this scenario, fallout tuffs at S2 and S3 accumulated above the sloping delta front, partially filling the surface relief, while pumice siltstone at S4, S5 and S6 accumulated on the flatter surface behind the delta front, where the tephra was more completely reworked and mixed with silt by the fluvial system that fed and cut across the top of the delta (Fig. 2b).

### Associated fauna and palaeoecological context

Only 26 vertebrate specimens have been collected from MRD-VP-1 to date. They sample a variety of mammalian taxa (Supplementary Table 6), most of which are common in other eastern African hominin-bearing sites of comparable age, but do not allow conclusive palaeoenvironmental inferences. Still, the sedimentary facies at MRD-VP-1, together with botanical remains separated from the lake and delta sediments (see below), suggest habitats that are associated with the margins of a lake, a delta and the riparian zone upstream.

The  $\delta^{13}\text{C}_{28\text{acid}}$  values of *n*-alkanoic acids derived from plant wax and pollen markers for Somalia-Masai shrubland/grassland (Poaceae, Amaranthaceae, *Indigofera* and Capparidaceae) in the same sediment samples indicate that the catchment for this basin was dry C<sub>3</sub>-dominated shrubland with some C<sub>4</sub> vegetation (Fig. 2c and Supplementary Tables 7–9). Only one sample was consistent with C<sub>4</sub> grass dominance ( $\delta^{13}\text{C}_{28\text{acid}} = -22.0\text{\textperthousand}$ ). Pollen abundances were insufficient for rigorous plant community analysis except for one sample, WM18-MRD-210, from the deltaic sandstone of MRD-VP-1 ( $n = 222$ ). For this sample, the habitat scores are highest for steppe (25%) and woodland (3%) habitats (Fig. 2d). On the basis of the high proportions of *Acacia*, Poaceae, and *Aerva*-type *javanica* in this sample (Extended Data Figs. 8, 9), we conclude that the catchment area for the delta was typical of dry, open *Acacia* woodland. The relatively high habitat scores for riparian forest (15%) and evergreen bushland (14%) types suggest that such environments were present upstream of, and close to, the hominin site. This sample (WM18-MRD-210) from the hominin-bearing sediment of MRD-VP-1 also yielded the lowest  $\delta^{13}\text{C}_{28\text{acid}}$  values ( $-30.6\text{\textperthousand}$ ) consistent with a greater proportion of C<sub>3</sub> trees, wetter or more closed canopy. We note that mature palaeosols with distinct horizons are not present in the lacustrine or deltaic strata, consistent with the high sediment accumulation rates and short exposure time for any land surfaces.

The evidence from the botanical proxies is broadly consistent with the wind-blown and regionally averaged signals recorded in marine sediments in the Gulf of Aden, for which the catchment area includes the Afar<sup>28</sup>. Specifically, the  $\delta^{13}\text{C}_{28\text{acid}}$  ( $-30.6$  to  $-22.0\text{\textperthousand}$ ;  $n = 15$ ) and  $\delta\text{D}_{28\text{acid}}$  ( $-100$  to  $-198\text{\textperthousand}$ ;  $n = 10$ ) values from the studied plant wax encompass the spatially averaged  $\delta^{13}\text{C}_{\text{wax}}$  ( $-28.9$  to  $-24.7\text{\textperthousand}$ ;  $n = 75$ ) and  $\delta\text{D}_{\text{wax}}$  ( $-146$  to  $-132\text{\textperthousand}$ ;  $n = 6$ ) results for the corresponding interval of a Gulf of Aden marine core<sup>29</sup>. Pollen from this core interval reveal that

modern-analogue assemblages, including variations in 10–50% Amaranthaceae and 9–30% grass<sup>30</sup>, were prevalent regionally at this time, indicating that the wetter conditions represented by the lake and delta sediments were local to this setting rather than products of a wet climatic interval.

The lake was at least 9 km<sup>2</sup> (based on exposures) and at least 6–8 m deep (based on clinoform heights). Alteration of volcanic glass combined with poor preservation of biogenic silica suggests alkaline conditions. In addition, LHG-248, the only sample to yield identifiable silica remains (Extended Data Fig. 9), is composed of 17 diatom species (10 genera), most of which are indicators of saline environments, predominantly (70%) those observed in coastal and marine environments (Supplementary Table 10). Furthermore, pink pigmentation in the lipid extract of some sediment samples is consistent with carotenoids in saline lakes<sup>31</sup>. The lake probably had a closed outlet at times. Diverse conglomerate clasts could have been derived from underlying Oligocene and Miocene basalts and rhyolites that were tilted steeply on west-southwest-dipping normal faults before approximately 4 Myr ago<sup>32</sup> (Extended Data Fig. 10). However, Afromontane taxa (*Hagenia abyssinica*) that today occur in the upper limit of the mountain forest<sup>33</sup> above 2,450 m are evidence that the catchment area of the lake included the Ethiopian highlands (Supplementary Table 9). This low-dispersal pollen may have been delivered by a river that—similar to the modern Mille River—descended the western margin of the Afar and turned southeastward around the palaeolandscape (Extended Data Fig. 10). The Gilbert-style delta and debris flows are features that are typical of high topographic relief and, together with the high sediment accumulation rates, are compatible with rapid filling of the space that was created by the crustal thinning<sup>32</sup> more than 4 Myr ago or by syn-depositional faulting. Although we have identified no faults that were demonstrably active during the time of the lake and delta, high-angle faults that offset the mapped strata and the eastward thickening of strata from S6 to S1 are evidence of continued tectonic activity. More work is needed to determine if, and how, lake deposits in the Godaya Valley relate to ≥3.8 Ma lake deposits elsewhere, especially along the Mille and Waki rivers, in southwestern Woranso-Mille, and in the Awash Valley (Extended Data Fig. 10). Although lakes are common in rift basins, it is notable that similar lake and Gilbert-type delta deposits of this age have not been documented elsewhere in the Afar. In general, the thick lacustrine strata of the Godaya Valley are not predicted by models of the Hadar Basin as an east-dipping half-graben, which place the thickest, deepest deposits near

inferred bounding fault(s) that are more than 40 km to the southeast, but are compatible with models of basin formation associated with rift-axis volcanic centres related to the Red Sea Rift<sup>9,12,34</sup>. In either case, the decreased sediment accumulation rate and increased alluvial conglomerate above the Mille tuff sequence and Sidi Hakoma tuff, respectively (Fig. 1b), indicate changes in basin dynamics and the landscape between the setting of the 3.80-Myr-old MRD-VP-1/1 hominin and younger fossils that have been recovered at other Woranso-Mille localities<sup>2–6</sup>.

**Online content** Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at

Received 10 April 2019; accepted 26 July 2019

<jrn>1. Deino, A. L. et al. <sup>40</sup>Ar/<sup>39</sup>Ar dating, paleomagnetism, and tephrochemistry of Pliocene strata of the hominid-bearing Woranso-Mille area, west-central Afar Rift, Ethiopia. *J. Hum. Evol.* **58**, 111–126 (2010). </jrn>

<jrn>2. Haile-Selassie, Y. et al. An early *Australopithecus afarensis* postcranium from Woranso-Mille, Ethiopia. *Proc. Natl Acad. Sci. USA* **107**, 12121–12126 (2010). </jrn>

<jrn>3. Haile-Selassie, Y. et al. A new hominin foot from Ethiopia shows multiple Pliocene bipedal adaptations. *Nature* **483**, 565–569 (2012). </jrn>

<jrn>4. Haile-Selassie, Y. et al. New species from Ethiopia further expands Middle Pliocene hominin diversity. *Nature* **521**, 483–488 (2015). </jrn>

<jrn>5. Haile-Selassie, Y. et al. Dentognathic remains of *Australopithecus afarensis* from Nefuraytu (Woranso-Mille, Ethiopia): comparative description, geology, and paleoecological context. *J. Hum. Evol.* **100**, 35–53 (2016). </jrn>

<jrn>6. Haile-Selassie, Y., Saylor, B. Z., Deino, A., Alene, M. & Latimer, B. M. New hominid fossils from Woranso-Mille (Central Afar, Ethiopia) and taxonomy of early *Australopithecus*. *Am. J. Phys. Anthropol.* **141**, 406–417 (2010). </jrn>

<jrn>7. Haile-Selassie, Y. et al. A 3.8-million-year-old hominin cranium from Woranso-Mille, Ethiopia. *Nature* (2019). </jrn>

- <jrn>8. Taieb, M., Johanson, D. C., Coppens, Y. & Aronson, J. L. Geological and paleontological background of Hadar hominid site, Afar, Ethiopia. *Nature* **260**, 289–293 (1976). </jrn>
- <edb>9. Campisano, C. J. & Feibel, C. S. in *The Geology of Early Humans in the Horn of Africa* Vol. 446 (eds Quade, J. & Wynn, J.) 179–201 (Geological Society of America, 2008). </edb>
- <edb>10. Wynn, J. G. et al. in *The Geology of Early Humans in the Horn of Africa* Vol. 446 (eds Quade, J. & Wynn, J.) 87–118 (Geological Society of America, 2008). </edb>
- <jrn>11. Renne, P. R., WoldeGabriel, G., Hart, W. K., Heiken, G. & White, T. D. Chronostratigraphy of the Miocene–Pliocene Sagantole Formation, Middle Awash Valley, Afar rift, Ethiopia. *Geol. Soc. Am. Bull.* **111**, 869–885 (1999). </jrn>
- <jrn>12. White, T. D. et al. Asa Issie, Aramis and the origin of *Australopithecus*. *Nature* **440**, 883–889 (2006). </jrn>
- <edb>13. Quade, J. et al. in *The Geology of Early Humans in the Horn of Africa* Vol. 446 (eds Quade, J. & Wynn, J.) 1–31 (Geological Society of America, 2008). </edb>
- <jrn>14. Asfaw, B. The Belohdelie frontal: new evidence of early hominid cranial morphology from the Afar of Ethiopia. *J. Hum. Evol.* **16**, 611–624 (1987). </jrn>
- <jrn>15. White, T. D. et al. New discoveries of *Australopithecus* at Maka in Ethiopia. *Nature* **366**, 261–265 (1993). </jrn>
- <jrn>16. Leakey, M. G., Feibel, C. S., McDougall, I. & Walker, A. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571 (1995). </jrn>
- <jrn>17. Leakey, M. G., Feibel, C. S., McDougall, I., Ward, C. & Walker, A. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* **393**, 62–66 (1998). </jrn>
- <edb>18. Deino, A. in *Paleontology and Geology of Laetoli: Human Evolution in Context* (ed. Harrison, T.) 77–97 (Springer, 2011).

- <jrn>19. Fleagle, J. G., Rasmussen, D. T., Yirga, S., Bown, T. M. & Grine, F. E. New hominid fossils from Fejej, southern Ethiopia. *J. Hum. Evol.* **21**, 145–152 (1991). </jrn>
- <jrn>20. Kappelman, J. et al. Age of *Australopithecus afarensis* from Fejej, Ethiopia. *J. Hum. Evol.* **30**, 139–146 (1996). </jrn>
- <jrn>21. Alene, M. et al. Geochemistry of Woranso-Mille Pliocene basalts from west-central Afar, Ethiopia: implications for mantle source characteristics and rift evolution. *Lithos* **282–283**, 187–200 (2017). </jrn>
- <jrn>22. Saylor, B. Z. et al. Tephrostratigraphy of the Waki-Mille area of the Woranso-Mille paleoanthropological research project, Afar, Ethiopia. *J. Hum. Evol.* **93**, 25–45 (2016). </jrn>
- <jrn>23. Walter, R. C. & Aronson, J. L. Age and source of the Sidi Hakoma Tuff, Hadar Formation, Ethiopia. *J. Hum. Evol.* **25**, 229–240 (1993). </jrn>
- <edb>24. Lourens, L., Hilgen, F., Laskar, J. & Wilson, D. S. in *A Geological Time Scale* (eds Gradstein, F. M. et al.) 409–440 (Cambridge Univ. Press, 2004). </edb>
- <edb>25. Saylor, B. Z. et al. in *The Postcranial Anatomy of Australopithecus afarensis* (eds Haile-Selassie, Y. & Su, D.) 13–23 (Springer, 2016). </edb>
- <edb>26. Middleton, G. V. & Hampton, M. A. in *Marine Sediment Transport and Environmental Management* (eds Stanley, D. J. & Swift, P.) 197–218 (Wiley, 1976). </edb>
- <edb>27. Elliot, T. in *Sedimentary Environments and Facies* (ed. Reading, H. G.) 113–154 (Blackwell Scientific, 1986). </edb>
- <jrn>28. Prospero, J. M. et al. Environmental characterization of global sources of atmospheric soil dust identified with the NIMBUS 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. *Rev. Geophys.* **40**, 1002 (2002). </jrn>
- <jrn>29. Liddy, H. M., Feakins, S. J. & Tierney, J. E. Cooling and drying in northeast Africa across the Pliocene. *Earth Planet. Sci. Lett.* **449**, 430–438 (2016). </jrn>
- <jrn>30. Bonnefille, R. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Global Planet. Change* **72**, 390–411 (2010). </jrn>

- <jrn>31. Winters, Y. D., Lowenstein, T. K. & Timofeeff, M. N. Identification of carotenoids in ancient salt from Death Valley, Saline Valley, and Searles Lake, California, using laser Raman spectroscopy. *Astrobiology* **13**, 1065–1080 (2013). </jrn>
- <jrn>32. Stab, M. et al. Modes of rifting in magma-rich settings: tectono-magmatic evolution of Central Afar. *Tectonics* **35**, 2–38 (2016). </jrn>
- <bok>33. Friis, I., Demissew, S. & van Breugel, P. *Atlas of the Potential Vegetation of Ethiopia* (The Royal Danish Academy of Sciences and Letters, 2010). </bok>
- <jrn>34. Lahitte, P., Gillot, P. Y. & Courtillot, V. Silicic central volcanoes as precursors to rift propagation: the Afar case. *Earth Planet. Sci. Lett.* **207**, 103–116 (2003). </jrn>
- <jrn>42. Prentice, C., Guiot, J., Huntley, B., Jolly, D., & Cheddadi, R. Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Clim. Dyn.* **12**, 185–194 (1996). </jrn>

[Author: Please ensure that the following information is included in the figure legends where relevant. Sample size (exact  $n$  number); a statement of replicability (how many times was experiment replicated in the laboratory); description of sample collection (clarify whether technical or biological replicates and include how many animals, litters, cultures, etc.); state the statistical test used and give  $P$  values; define centre values (median or average) and error bars. For figures/images that are reproduced or adapted from a third party, it is important that you confirm that permission has been obtained and that appropriate acknowledgement of the copyright holder is given.]

**Fig. 1 | Stratigraphic setting of the fossil cranium.** **a**, The location of the Godaya Valley relative to fossiliferous areas near the Waki-Mille confluence and east of Korsi Dora. **b–d**, Geological features of the Godaya Valley are illustrated as simplified stratigraphic columns (**b**), a partial geological map (**c**) and a cross-section along line of topographic profile with 10:1 vertical exaggeration (**d**). **b**, White and black rectangles next to section S6 indicate reverse and normal palaeomagnetic zones that correlate with the geopolarity time scale (GPTS)<sup>24</sup>. **c, d**, Ages of tuff correlates are from previous studies<sup>1,22,25</sup> **d**, Colours of mapped units are as in **b**; uncoloured areas are younger cover or not mapped; the documented stratigraphic offset is a criterion for fault identification in **c**. Note that changes in apparent dip in the cross-section

correspond to changes in the orientation of the topographic profile. m a.s.l., metres above sealevel. Supporting data are reported in Supplementary Tables 1–4. Data for the base map and topographic profile are from Google, DigitalGlobe.

**Fig. 2 | Lake, delta and tuffaceous strata below and above the level where the hominin cranium was found.** **a**, Measured stratigraphic sections below the Am-Ado basalt group hung on the dated T9 tuff. **b**, Cartoon illustrating the southeastward progradation of the delta front (thick black lines) over the lake bed tuffs (T1–T8) and the subsequent partial infill of topography on top of the delta by overlying tuffaceous strata (T9–Mille tuff sequence). Rose diagram shows palaeocurrent directions measured on trough axes and foreset dips in the deltaic sandstone. **c**, Stable hydrogen and carbon isotope data from *n*-alkanoic acids in claystone, siltstone and sandstone samples ( $n = 16$ ) plotted in their relative stratigraphic positions. Error bars based on the s.d. of instrument duplicates are smaller than the size of the symbols, with the exception of the  $\delta^{13}\text{C}$  value above T5. **d**, Habitat scores for sample WM18-MRD-210 calculated based on the percentage of pollen taxa with affinity for each habitat<sup>42</sup>. EB, evergreen bushland; MF, montane forest; RF, riparian forest; Sw, swamp. Supporting data are reported in Supplementary Tables 2–10.

[Author: Please ensure that the following information is provided in the Methods section where relevant. We recommend that detailed protocols are deposited in Protocol Exchange, or a similar repository. If custom computer code has been used and is central to the conclusions of this paper, please insert a section into the Methods titled ‘Code availability’ and indicate within this section whether and how the code can be accessed, including any restrictions to access.]

## METHODS

### Basalt geochemistry

Powdered basalt samples were analysed for major and trace element concentration using a PANalytical 2404 X-ray fluorescence vacuum spectrometer at Franklin and Marshall College, following previously described techniques<sup>35</sup>. This includes determination of ferrous iron by standard titration methods and total volatile content by loss on ignition. Samples were assigned to chronostratigraphically distinct basalt groups based on the similarity of element concentrations and stratigraphic position to previously characterized basalt groups<sup>21</sup> (Supplementary Table 1).

### Tephra geochemistry

Tuff samples were prepared for glass geochemical analysis as previously described<sup>22</sup>. Elemental concentrations in felsic volcanic glass were measured in polished mounts using wavelength dispersive spectroscopy at the University of Wisconsin on Cameca SX-51 and Cameca SX-Five Fe electron probe micro-analyzers. The mounts were carbon-coated at the same time as standards, which included Kilbourne olivine for Mg, Minas Gerais rutile for Ti, Monash andesine for Ca, synthetic manganese olivine for Mn, Rockport fayalite for Fe, topaz for F, USMN chlorapatite for Cl and Lipari obsidian for Si, Al, O, Na and K. Analyses were conducted using an accelerating voltage of 15 KeV, a beam current of 10 nA, and a focused beam with onpeak and offpeak count times of 20 s. Element concentrations were converted to oxides computing all Fe as Fe<sub>2</sub>O<sub>3</sub>, corrected for beam-induced ion migration of NaO, O and K<sub>2</sub>O using time-dependent intensity data, and normalized for differences in water content. Analyses with totals less than 97.5 or greater than 102.5 were excluded, except for MRD-17-4, for which analyses with water-adjusted totals less than 89.5 were excluded. The normalized oxide composition was compared with chemically characterized tuffs near the Waki-Mille confluence<sup>22</sup> (Supplementary Table 2).

### <sup>40</sup>Ar/<sup>39</sup>Ar dating

Two fallout tuff horizons were analysed using the single-crystal <sup>40</sup>Ar/<sup>39</sup>Ar laser incremental-heating dating method. These samples were collected in stratigraphic sequence at Lahaysule Gera hill (S2); sample WM18/LHG-1 is from a 2-cm thick medium-to-fine-grained tuff (30 cm below tuff unit T0) at the base of the exposures in the lacustrine sequence that underlies the area, while WM12/LHG-6 was obtained around 2 m above the base of a prominent, approximately 3 m thick, coarsely crystalline white lapillistone tuff (unit T9) above the lacustrine beds as well as deltaic siltstone. Following standard mineral purification procedures, WM18/LHG-1 yielded K-feldspar phenocrysts for analysis (approximately 0.2–0.4-mm grain size), whereas WM12/LHG-6 lacked K-feldspar but yielded plagioclase samples (around 0.45–1.5 mm) that were suitable for analysis. The plagioclase separate from WM12/LHG-6 was irradiated for 0.5 h in the Cd-lined in-core CLICIT facility of the Oregon State University TRIGA reactor. Two laboratory identifiers were assigned to this material (26236 and 26238), representing near-adjacent aliquot positions around the ring of a single aluminium irradiation sample holder. K-feldspar from WM18/LHG-1 was irradiated in the same facility for 10 h. Both irradiations used

sanidine phenocrysts from the Alder Creek Rhyolite of California as the monitor mineral (age =  $1.1848 \pm 0.006$  Myr)<sup>36</sup>. Reactor-induced isotopic production ratios for this irradiation were:  $(^{36}\text{Ar}/^{37}\text{Ar})_{\text{Ca}} = 2.65 \pm 0.02 \times 10^{-4}$ ,  $(^{38}\text{Ar}/^{37}\text{Ar})_{\text{Ca}} = 1.96 \pm 0.08 \times 10^{-5}$ ,  $(^{39}\text{Ar}/^{37}\text{Ar})_{\text{Ca}} = 6.95 \pm 0.09 \times 10^{-4}$ ,  $(^{37}\text{Ar}/^{39}\text{Ar})_{\text{K}} = 2.24 \pm 0.16 \times 10^{-4}$ ,  $(^{38}\text{Ar}/^{39}\text{Ar})_{\text{K}} = 1.220 \pm 0.003 \times 10^{-2}$ ,  $(^{40}\text{Ar}/^{39}\text{Ar})_{\text{K}} = 2.5 \pm 0.9 \times 10^{-4}$ . Atmospheric  $^{40}\text{Ar}/^{36}\text{Ar} = 298.56 \pm 0.31$  from ref. <sup>37</sup> and decay constants were according to ref. <sup>38</sup>.

Following irradiation, phenocrysts were analysed individually ( $n = 13$  for WM12/LHG-6;  $n = 21$  for WM18/LHG-1) under ultra-high vacuum using a CO<sub>2</sub> laser with top-hat beam profile at a diameter of 2.3 mm. Argon analyses were executed on-line with the extraction system, using a five-detector Nu Instruments Noblesse mass spectrometer operating in ion-counting, simultaneous collection mode. All grains of WM12/LHG-6 were analysed incrementally from a very low degas temperature to fusion in six steps (Extended Data Fig. 4 and Supplementary Table 3). However, in the case of WM18/LHG-1, each grain was analysed in one or two low-power steps to allow a preliminary assessment of age and chemistry; grains that were too old (more than 10 Myr old,  $n = 12$ ) or exhibited a Ca/K ratio of plagioclase ( $n = 1$ ) were omitted, leaving 8 grains taken to completion of the incremental heating sequence. Further details of irradiation procedure, argon analysis and data reduction are provided in a previously published study<sup>24</sup>.

#### Palaeomagnetic stratigraphy

Three main areas were selected, measured and sampled to analyse the magnetostratigraphy. A stratigraphic section that included the lowest and highest outcropping strata in the local area was produced for each of these three areas, which are, from west to east: Guda plateau (S6), the area around MRD-VP-1 (S4), and Lehaysule Gera hill (S2), northeast of the Am-Ado plateau. The longest section (S6) is a composite of two sections measured on both sides of a gully. Section S4 extends from the lowest strata below the fossiliferous sandstones to the top of the Miro Dora plateau, including tuffs and sediments above the Am-Ado basalts in the northwestern portion of the plateau. In total, 33 samples were collected and analysed from these areas: 24 at S6 (part a and b), 5 at S4 and 4 at S2. Samples were collected from fresh outcrops after digging and removing weathered superficial material. Mainly fine-grained sediments (clay and silt) were selected for sampling. Block samples of approximately 500 cm<sup>3</sup> were cut using manual tools and

one flat face was oriented using a Brunton compass. In the laboratory, each sample was sawed, sanded and cleaned with compressed air and separated into at least 3 cubic specimens ( $10.5\text{ cm}^3$ ). Measurements were done with a 2G three-axis cryogenic magnetometer at the University of Barcelona/CSIC palaeomagnetic laboratory. At least 2 specimens from each sample (a total of 79 specimens) were thermally demagnetized in a non-inductive furnace (less than 10 nT), at 50 °C steps up to 650 °C to remove secondary magnetizations and to isolate the primary characteristic magnetic direction. Magnetic susceptibility was measured with a KLY02 susceptibility bridge (Agico) at each demagnetization step to monitor possible mineralogical changes during the heating process.

Samples in general are of good quality and most show non-ambiguous palaeomagnetic directions with all specimens from each sample showing consistent normal or reverse directions. In a limited number of specimens, the secondary magnetization was not completely removed showing ambiguous directions. The magnetic susceptibility was measured after each thermal demagnetization step, primarily to monitor mineralogical changes that might occur during heating. Samples show high magnetic susceptibility and no important changes are shown until the sample was heated to high temperatures of more than 500 °C. Overall, the samples yielded unblocking temperatures of more than 600 °C, indicating that magnetite was the main carrier of the magnetization. Palaeomagnetic directions were calculated using principal component analysis with the Paldi software (Extended Data Fig. 7 and Supplementary Table 4).

In general, samples show a low temperature component of the magnetization that was removed after heating at 250–300 °C, allowing isolation of a primary high temperature component. In some specimens, the secondary component of the magnetization was not completely removed after demagnetization, and the resulting remnant directions are a mixture of normal and reverse polarities. These specimens show ambiguous directions, such as declinations towards the south, but positive inclinations or declinations towards the north with a negative inclination. We note also that sample MRD17 shows evidence of weathering with formation of new iron-rich minerals and a suspicious normal magnetization between two reverse samples. The 58 non-ambiguous specimens were selected to calculate mean normal and reverse characteristic remnant magnetizations and build a local polarity sequence correlated with the geopolarity time scale<sup>24</sup> based on available  $^{40}\text{Ar}/^{39}\text{Ar}$  dates (Extended Data Fig. 7).

**Leaf wax extraction and isotopic analysis**

Sediment samples including claystone ( $n = 8$ ), siltstone ( $n = 7$ ) and sandstone ( $n = 1$ ) were analysed for leaf-wax compound-specific isotopes (Supplementary Tables 7, 8). Samples were fragmented with a hammer before powdering with a ring and puck mill. Lipids were extracted using a Dionex Accelerated Solvent system with 9:1 dichloromethane (DCM):methanol at 100 °C and 1,500 psi. The total lipid extract was separated into neutral and acid fractions over NH<sub>2</sub> sepra bulk packing, the neutral fraction was eluted using 2:1 DCM:isopropanol and the acid fraction was eluted with 4% formic acid in diethyl ether. The acid fraction was methylated in a mixture of 95:5 methanol:hydrochloric acid at 70 °C for 12 h, using methanol of known isotopic composition ( $\delta^{13}\text{C} = -24.7\text{\textperthousand}$  and  $\delta\text{D} = -187\text{\textperthousand}$ ). Excess Milli-Q water was added and the fatty acid methyl esters (FAMEs) were partitioned into hexane, dried by passing through anhydrous Na<sub>2</sub>SO<sub>4</sub> and purified over a silica gel column and eluted with DCM. Unsaturated compounds were removed from samples by passing over a silver-nitrate-treated silica gel column, eluting the saturated FAMEs with DCM.

*n*-Alkanoic acids were identified using an Agilent 6890 gas chromatograph equipped with an Rxi-5ms (30m × 0.25mm, film thickness 0.25 μm) column connected to a 5973 MSD mass spectrometer, and quantified using the flame ionization detector. Samples were dissolved in 500 μL of hexane and one microlitre of this solution was injected into the inlet using a 7683 programmable autosampler. Quantification of FAMEs was conducted through comparison of sample peak areas in chromatograms to peak areas of an in-house standard containing known concentrations of *n*-alkanoic acids.

Carbon and hydrogen isotopes of *n*-alkanoic acids were measured using a Thermo Scientific Trace gas chromatograph equipped with a Rxi-5ms column (30 m × 0.25 mm; film thickness, 1 μm) with a PTV injector operated in solvent-split mode, coupled to a Delta V Plus isotope ratio mass spectrometer using an Isolink combustion/pyrolysis furnace (1,000/1,420 °C). A standard of known isotopic compositions (A6 mix supplied by A. Schimmelmann, University of Indiana) was measured daily, allowing for normalization to the Vienna Standard Mean Ocean Water/Standard Light Antarctic Precipitation and Vienna Pee Dee Belemnite/Lithium carbonate standard prepared by L. Svec isotopic scales for δD and δ<sup>13</sup>C values, respectively. The root mean square error of the replicate analyses of the external standard was 4‰ for δD and 0.1‰ for δ<sup>13</sup>C.

Sample  $\delta^{13}\text{C}$  and  $\delta\text{D}$  values were corrected using a mass balance for the addition of the methyl group of known isotopic composition as previously described<sup>39</sup>.

#### **Pollen extraction and analysis**

We prepared 16 samples—most of which were the same as those analysed for leaf wax isotopes and phytoliths (Supplementary Table 7)—for microscopic pollen analysis following a standard procedure: (1) samples (5–30 g, depending on the sediment that was available) were immersed in HCl (33%, 1 h) to remove carbonates; (2) digestion of silicate minerals in HF (48%, 12 h); (3) immersion in HCl (33%, 4 h) to remove fluorosilicates produced in the previous step; (4) immersion in KOH (20%, 10 minutes in a hot water bath) to remove organic matter; (5) sieving at 150  $\mu\text{m}$ ; (6) heavy liquid separation using sodium polytungstate set at a specific gravity of 2.2 *grams per cm<sup>3</sup>* to improve pollen concentration. For productive samples, the entire residue sample was analysed under a microscope. Slides were prepared with glycerin and sealed with Entellan. They were scanned at a magnification of  $\times 250$  and identified at a magnification of  $\times 1,000$ .

Pollen identifications (Supplementary Table 9) are based on the pollen reference collection available at CEREGE, Aix-en-Provence and published pollen atlases.

Palaeovegetation inferences are based on modern pollen studies from east as well as west Africa<sup>4</sup>, botanical surveys, interpretations proposed earlier by R. Bonnefille for Hadar and Middle Awash<sup>30,40,41</sup>, and unpublished modern pollen data of B.B. from the Afar region (Supplementary Table 9). For the MRD-210 pollen assemblage, habitat scores were calculated based on pollen taxa diversity and pollen percentages following the biomization procedure<sup>42</sup>. This method allows a given taxon to be assigned to different habitats. Pollen percentages are square-rooted to reduce the weight of overrepresented taxa. The greater the number of taxa contributing to a given habitat and the higher their relative abundances (as a percentage), the higher the score of the given habitat is. Consequently, habitat types with higher scores can be considered closer (because these are best represented in terms of taxa diversity and abundance) than those with lower scores. Poaceae were excluded from score calculation to improve the signal.

#### Biogenic silica extraction and analysis

The same 16 samples that were analysed for leaf wax isotopes were also processed for phytoliths using a process modified from a previous study<sup>43</sup>. The processing was done on approximately 1 g of the original material for each sample. Hydrochloric acid and nitric acid (heated) steps were, respectively, used to remove the carbonates and oxidize the organic matter. They were alternated with two sieving steps combined with numerous rounds of centrifugation to first remove the largest mineral fraction ( $\geq 250 \mu\text{m}$ ), and then the smallest fraction, which included clays. Finally, a densimetric separation of the biogenic silica fraction using a zinc bromide heavy liquid set at a  $d = 2.3$  was performed. Residues were obtained in minor amounts for all samples. Owing to their scarcity, samples were all mounted on slides using Canada balsam medium and observed at  $\times 1,000$  magnification. Sample WM18-MRD-210 is the only exception with abundant residues; however, these were mostly volcanic ash particles rather than biogenic silica. WM18-LHG-248, the only sample to yield identifiable taxa, including diatoms, was reprocessed with 2–3 g of material to enable species identification (Supplementary Table 10).

#### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

#### Data availability

Any data supporting the findings of this study that are not included in the paper and Supplementary Information are available from the corresponding author upon reasonable request.

#### Code availability

A data analysis version of the  $^{40}\text{Ar}/^{39}\text{Ar}$  dating software is available at no cost from A.D. (adeino@bge.org). The Bayesian age–stratigraphic analysis model is available at <https://github.com/brehinkeller>.

<jrn>35. Mertzman, S. A. K–Ar results from the southern Oregon–northern California Cascade range. *Or. Geol.* **62**, 99–122 (2000).</jrn>

- <jrn>36. Niespolo, E. M., Rutte, D., Deino, A. L. & Renne, P. R. Intercalibration and age of the Alder Creek sanidine  $^{40}\text{Ar}/^{39}\text{Ar}$  standard. *Quat. Geochronol.* **39**, 205–213 (2017).</jrn>
- <jrn>37. Lee, J.-Y. A redetermination of the isotopic abundances of atmospheric Ar. *Geochim. Cosmochim. Acta* **70**, 4507–4512 (2006).</jrn>
- <jrn>38. Min, K. W., Mundil, R., Renne, P. R. & Ludwig, K. R. A test for systematic errors in  $^{40}\text{Ar}/^{39}\text{Ar}$ -geochronology through comparison with U/Pb analysis of a 1.1-Ga rhyolite. *Geochim. Cosmochim. Acta* **64**, 73–98 (2000).</jrn>
- <jrn>39. Lee, H. et al. Comparison of three methods for the methylation of aliphatic and aromatic compounds. *Rapid Commun. Mass Spectrom.* **31**, 1633–1640 (2017).</jrn>
- <jrn>40. Bonnefille, R., Vincens, A. & Buchet, A. Palynology, stratigraphy and palaeoenvironment of a Pliocene hominid site (2.9–3.3 M.Y.) at Hadar, Ethiopia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **60**, 249–281 (1987).</jrn>
- <jrn>41. Bonnefille, R., Potts, R., Chalié, F., Jolly, D. & Peyron, O. High-resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. *Proc. Natl Acad. Sci. USA* **101**, 12125–12129 (2004).</jrn>
42. Prentice, C., Guiot, J., Huntley, B., Jolly, D., & Cheddadi, R. Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Clim. Dyn.* **12**, 185–194 (1996).
- <ths>43. Strömberg, C. A. E. The Origin and Spread of Grass-Dominated Ecosystems during the Tertiary of North America and How It Relates to the Evolution of Hypsodonty in Equids. PhD Thesis, Univ. California Berkeley (2003).</ths>
- <jrn>44. Deino, A. L. et al. Chronostratigraphic model of a high-resolution drill core record of the past million years from the Koora Basin, south Kenya Rift: overcoming the difficulties of variable sedimentation rate and hiatuses. *Quat. Sci. Rev.* **215**, 213–231 (2019).</jrn>
- <unknown>45. Miall, A.D. *The Geology of Fluvial Deposits: Sedimentary Facies, Basin Analysis and Petroleum Geology* (Springer-Verlag, 1996).</unknown>

- <jrn>46. DiMaggio, E. N. et al. Late Pliocene fossiliferous sedimentary record and the environmental context of early *Homo* from Afar, Ethiopia. *Science* **347**, 1355–1359 (2015). </jrn>
- <jrn>47. Kirschvink, J. L. The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. Int.* **62**, 699–718 (1980). </jrn>
- <edb>48. Sournia, A. *Diatomées planctoniques du Canal de Mozambique et de l'Île de la Réunion*. 1-120 (ORSTOM, 1968). </edb>
- <unknown>49. Foged, N. *Some Littoral Diatoms from the Coast of Tanzania*. 1-127 (*Biblioteca Phycologia*, 1975).
- <jrn>50. Ryan, W. B. F. et al. Global multi-resolution topography synthesis. *Geochem. Geophys. Geosyst.* **10**, Q03014 (2009). </jrn>

**Acknowledgements** We thank the Authority for Research and Conservation of Cultural Heritage for permission to conduct field and laboratory work, the Afar people of Woranso-Mille for their hospitality, and the project's fieldwork crew members for their tireless support of field activities; M. Atkins and I. Van Orman for assistance with the preparation of tephrochemical samples and J. Fournelle for assistance with their analysis. This research was supported by grants from the Spanish Government CGL2016-79458-P, Catalan Government 2017-SGR 824, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 659596 and the US National Science Foundation (1124705, 1124716, 1125157, 1125345, 1322017).

**Author contributions** B.Z.S., L.G., M.A., A.D., N.E.L., S.M.M., M.D.P. and Y.H.-S. conducted field research. M.A. collected basalt samples and S.A.M. analysed the samples for major and trace element concentrations. B.Z.S. collected and analysed tuff samples for major and minor element abundances in glass. A.D. collected and analysed tuff samples for  $^{40}\text{Ar}/^{39}\text{Ar}$  dating. L.G. collected sediment samples and analysed their palaeomagnetic polarity. N.E.L. and M.D.P. collected sediment samples for analysis of botanical remains. S.M.M. contributed to the collection and analysis of mammalian fossils. M.D.P. and S.J.F. analysed and interpreted leaf wax isotopes. B.B. and D.B. separated and identified pollen. A.N. and F.S. separated and identified phytolith and diatom components. Y.H.-S. collected and analysed the mammalian fossils and directed the field station. B.Z.S. took the lead in writing the paper with contributions from all co-authors.

**Competing interests:** The authors declare no competing interests.

Additional information

**Supplementary information** is available for this paper at

**Correspondence and requests for materials** should be addressed to Y.H.-S.

**Peer review information** *Nature* thanks Craig S. Feibel, John W. Kappelman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

[Author: Please ensure that the following information is included in the figure legends where relevant. Sample size (exact  $n$  number); a statement of replicability (how many times was experiment replicated in the laboratory); description of sample collection (clarify whether technical or biological replicates and include how many animals, litters, cultures, etc.); state the statistical test used and give  $P$  values; define centre values (median or average) and error bars. For figures/images that are reproduced or adapted from a third party, it is important that you confirm that permission has been obtained and that appropriate acknowledgement of the copyright holder is given.]

**Extended Data Fig. 1 | Field photographs.** **a**, Location of the horizon of the MRD-VP-1 fossil relative to features on the Miro Dora plateau (capped by Am-Ado basalts) and the higher Guda plateau (background). **b**, Coarse sand matrix on MRD-VP-1/1. Additional matrix types, which are not shown, included very hard fine sand and soft silty clay. **c**, Lehaysule Gera hill (section S2); image shows the position of geochemically identified tuffs (Mille tuff sequence and associated basaltic tuffs, and Araskimiro tuff) and the dated T9 tuff (WM12/LHG-6) above dark deltaic siltstone and lacustrine claystone. A lower dated tuff (WM18/LHG-1) near the bottom of the claystone exposure is not discernable. **d**, The hill of section S5; image shows the position of the deltaic sequence relative to the geochemically identified Mille tuff and associated basaltic tuff sequence and T9 tuff. **e**, Large-scale cross-bedded sandstone at MRD-VP-1. Arrows show foreset dip directions, indicating the flow toward the southeast (right arrow) and toward the south-southwest (left arrow). **f**, Upward-coarsening delta lobe(s) along the Am-Ado Plateau (S3) with Mille tuff sequence and Am-Ado basalt above and between.

**Extended Data Fig. 2 | Tephrochemical data and age–depth models.** **a**, Bivariate plots of oxide abundances in volcanic glass normalized for differences in water content (Supplementary Table 2). Symbol colours and shapes are as in Fig. 1 with closed, smaller symbols indicating newly reported analyses of individual glass shards from samples from the study area, differentiated by symbol border colour and size. Open symbols indicate previously published<sup>22</sup> sample means with  $1\sigma$  variation for named tuffs near the Waki-Mille confluence, specifically:

Kilaytoli tuff (WM-KSD-3,  $n = 19$ , 11 high Fe and 8 low Fe), Mesgid Dora tuff (MSD-08-5,  $n = 12$ ), Araskimiro tuff (MSD-08-3R,  $n = 12$  and WM-MD5 P,  $n = 10$ , duplicate samples from same location), bimodal tuff (WMC-08-3,  $n = 16$ ,  $n = 10$ , duplicate analyses of same sample), Mille tuff sequence (ARI-08-3,  $n = 16$ , ARI-08-4,  $n = 11$ , two chemically similar, vertically adjacent tuffs in the sequence). The Mille tuff sequence, the rhyolitic component of the basaltic and rhyolitic ‘bimodal tuff’ and their correlates are distinguished from other pumice tuffs in the area by their low CaO abundances. The Araskimiro tuff and its correlates are distinguished from chemically similar pumice tuffs in the area, including the Mesgid Dora tuff, by position above the Mille tuff/basaltic tuff sequence and below the Waki tuff, which is present below the Mesgid Dora tuff in a drainage approximately 300 m to the east of S2 (Fig. 1d). **b**, Bayesian age-stratigraphic model<sup>44</sup> for  $n = 3$  tuffs at S2, including the lowermost dated tuff (WM18/LHG-1), the T9 tuff (WM12/LHG-6) and the bottom of the Mille tuff sequence (around 2 m below LLG-12-10, which is chemically similar to the Mille/basaltic tuff sequence), and a similar model correlating these horizons to S6.

**Extended Data Fig. 3 | Detailed stratigraphic sections with positions of samples that were investigated and analysed for leaf wax, pollen and phytolith content.** All sections are in cm except for MRD18-03. Names of sections are listed at the top of each column and equivalent section names reported in Fig. 2 are listed in parentheses. The number associated with each sample ID is listed next to the dot that indicates the stratigraphic position of samples. If sample numbers are equivalent stratigraphically to other samples, this is noted in parentheses. Tuff numbering is included on the left of each column. Dashed lines provide ties between stratigraphic and lithologic equivalent horizons for non-tuffaceous units. The tuff numbering scheme also indicates tie points. Facies classification abbreviations are from a previously published study<sup>45</sup>. F, S and G indicate grain size, clay or silt size, sand size and gravel size, respectively, and the descriptors m, h, l, p and t describe the primary bedding as massive, horizontal, laminated, planar and trough, respectively. Listed latitudes and longitudes are in reference to the WGS-84 datum and indicate the position of the base of the section unless otherwise noted. The stratigraphic positions for samples WM18-MRD-210, -413, -414, -416 (section S4) and WM18-LHG-258 (section S2) can be found in Fig. 2.

**Extended Data Fig. 4 | Incremental heating release spectra from single-grain  $^{40}\text{Ar}/^{39}\text{Ar}$  analyses.** The x axis of each subplot is the cumulative percentage release of  $^{39}\text{Ar}$  during the

course of the experiment. ‘Int. Age’ is the ‘integrated’ or total-gas age, obtained by isotopically summing the individual steps weighted by the inverse variance, with the error calculated as the modified standard error at  $1\sigma$  (standard error of the weighted mean multiplied by the square root of the MSWD when  $\text{MSWD} > 1$ ). The apparent age plateau, defined as a sequence of at least three or more consecutive steps representing at least half of the total  $^{39}\text{Ar}$  release, and for which analytical error alone is sufficient to explain the observed age dispersion at a probability of 95%, is calculated as the mean age weighted by the inverse variance, with the error shown as the  $1\sigma$  modified standard error. Neither the integrated age nor plateau age incorporated the error in  $J$  at this stage in the data reduction; this systematic error is included in the final weighted-mean age calculated from the isochron results. Results for sample WM12/LHG-6 indicated  $n = 10$  plateaus identified out of a total of  $n = 13$  single-crystal incremental heating experiments; for sample WM18/LHG-1,  $n = 8$  plateaus were found in  $n = 8$  heating experiments.

**Extended Data Fig. 5 | Inverse-isochron correlation plots of plateau steps from incremental heating experiments.** Plots show isotope correlation diagrams for  $^{36}\text{Ar}/^{40}\text{Ar}$  and  $^{39}\text{Ar}/^{40}\text{Ar}$  ratios. Each grain is plotted separately. The age is obtained from the  $x$ -axis intercept and is shown with a  $1\sigma$  modified standard error. ‘ $^{40}\text{Ar}/^{39}\text{Ar}$  Int.’ refers to the ‘trapped’ non-radiogenic  $^{40}\text{Ar}/^{36}\text{Ar}$  ratio derived from the  $y$ -axis intercept of the isochron. ‘P’ refers to the probability that the observed scatter can be explained by analytical errors alone; a value below 0.05 indicates that analytical errors are insufficient to explain the magnitude of the dispersion (that is, geological scatter is present). ‘n’ indicates the included number of analyses.

**Extended Data Fig. 6 | Age-probability density plot of ages derived from isochron analyses.** **a**, Mean Ca/K atomic ratio of individual grains estimated from  $^{37}\text{Ar}/^{39}\text{Ar}$  measurements. **b**, Rank-order plot of single-crystal isochron results with  $1\sigma$  analytical uncertainty. **c**, The vertical scale is the relative likelihood of a given age occurring in the sample. The weighted-mean ages of the population are shown, the  $1\sigma$  error includes the error in  $J$ , the neutron-fluence parameter. The MSWD and number of included analyses are included after the mean age. Values in purple were omitted from analyses based on distance from the median age, with the cut-off level of 2.0 normalized median absolute deviations<sup>46</sup>. For sample WM12/LHG-6, 8 out of 10 isochron results are included in the weighted mean; for sample WM18/LHG-1, 6 out of 8 results are included.

**Extended Data Fig. 7 | Palaeomagnetic results.** **a**, Magnetostratigraphy for sections S6, S4 and S2 with results expressed as the angle to expected normal direction ( $\Delta$ ). The reverse–normal–reverse polarity sequence in S6 is correlated with the geopolarity time scale (GPTS)<sup>24</sup> on the basis of dated and geochemically identified tuffs in S2. White circles are mean directions of specimens (black), open circles are ambiguous specimen directions. vf = very fine sand, m = medium sand, C = coarse sand. **b**, Stereoplots and orthogonal demagnetization diagrams for a reverse (MRD-19a) and a normal (MRD-22b) sample. dec = declination, inc = inclination, int = intensity, NRM = natural remanent magnetization. The plots show the stepwise demagnetization process and the characteristic remnant magnetization directions (red lines) calculated using a principal component analysis<sup>47</sup>. In the stereonet plots (left) open circles show negative inclinations and solid dots show positive inclinations. In the orthogonal plots (right) solid dots indicate declination values and empty circles inclination values. Intensity units are in  $10 \text{ A m}^{-1}$ . mad, maximum angle deviation. **c**, Stereographic projection, referred to magnetic North of characteristic remnant magnetization directions for specimens collected at S6, S4 and S2. Circles indicate individual specimens. The plot shows the associated normal and reversed mean directions (squares) of the specimens, as determined by Fisher statistics, and the position of the present field (triangle). Only those samples showing non-ambiguous directions are plotted. In total, 48 specimens show reverse polarity and 10 samples show normal polarity.

**Extended Data Fig. 8 | Pollen diagrams for sample WM18-MRD-210 and diatom diagram for WM18-LHG-248.** **a**, Relative abundances of pollen taxa (%) in WM18-MRD-210 calculated using the total number of pollen grains and fern spores (\*; -type). Numbers in brackets refer to the number of pollen grains. **b**, Relative abundance of arboreal pollen (AP; trees and shrubs), non-arboreal pollen (NAP; herbs and forbs) and non-classified habitus (NC) in WM18-MRD-210. Numbers in brackets refer to the number of pollen grains. aqua., aquatic. **c**, Distribution of diatoms in WM18-LHG 248 according their ecological affinity: coastal–marine (yellow) versus brackish lacustrine (blue) environments. Among the species with coastal–marine affinities, *Coscinodiscus lineatus* and *Coscinodiscus excentricus* are typically observed in the Indian Ocean and Mozambique Channel<sup>48,49</sup>. Numbers in brackets refer to the number of diatom samples.

**Extended Data Fig. 9 | Examples of fossil pollen grains (sample WM18-MRD-210) and biogenic silica remains (sample WM18-LHG-248).** **a–d** Pollen grains. **a**, *Aerva*-type *javanica*.

**b**, *Euclea*. **c**, *Acacia*. **d**, *Tribulus*. **e–g**, Phytolith bodies. **e**, A spherical/globular psilate body, probably an indicator of dicot tree/shrub taxa. **f**, **g**, Two unidentified phytolith remains. **h**, A sponge spicule. **i–l**, Diatom frustules. **i**, *Cyclotella meneghiniana*. **j**, *Actinoptychus splendens*. **k**, *Denticula valida*. **l**, *Cyclotella stylorum*. Scale bars, 10 µm.

**Extended Data Fig. 10 | Shaded relief map of west central Afar.** Volcanic and sedimentary strata of the Sullu Adu Range, including the Guda Plateau, are sub-horizontal, 4 Myr old and younger, and stratigraphically above Oligocene and Miocene basalts and rhyolites that are tilted to the east-northeast, indicating periods of extension<sup>32</sup> before approximately 4 Myr ago. Nearby areas with well-studied exposures<sup>13,32</sup> of these Oligocene and Miocene formations are labelled in bold. Areas with lake deposits<sup>10,13</sup> in the lower Hadar or upper Sagantole formations are labelled in italics. The base map is from <http://www.geomapapp.org> using the global multi-resolution topography synthesis<sup>50</sup>.



## **Utilisation des pollens pour la reconstruction à haute résolution spatiale des environnements à Hominini du Plio-Pléistocène dans le rift est-africain (Éthiopie et Tanzanie)**

Cette thèse présente l'étude du signal pollinique de sédiments modernes et fossiles pour contribuer à une meilleure connaissance des micro-habitats associés aux Hominini du Plio-Pléistocène du rift africain. Grâce à un jeu de données modernes de 283 spectres polliniques de surface (dont 62 nouveaux) centré sur l'Ethiopie, il a été développé une approche qui permet maintenant d'identifier 24 types de végétations dont 9 ripariennes. Appliquée aux 42 spectres (dont 9 nouveaux) des Formations pliocènes de Hadar et Busidima (basse vallée de l'Awash, Éthiopie), cette approche montre que, entre 3,42 et 2,96 Ma, le Nord Afar a subi d'importants refroidissements (-6,5 à -10,5°C) associés aux événements globaux des stades isotopiques marins M2, MG2 et MG4. Ceux-ci ont induit d'importants changements de végétations à l'échelle régionale durant cette période, alors qu'au même moment des formations arborées se maintenaient, localement, à proximité de la paléorivière Awash et de résurgences d'eaux souterraines offrant ainsi un habitat boisé stable à *Australopithecus afarensis*. L'analyse pollinique d'un sondage sub-moderne en bordure du Lac d'Eyasi (Tanzanie) éclaire sur la dynamique de végétation associée aux résurgences et au climat régional en zone aride. Enfin, les reconstructions de paléovégétations associées aux Hominini étant limitées par les échantillons sub-stériles, il est développé ici un système de reconnaissance automatique. Basé sur des réseaux de neurones convolutifs, ce système est capable d'identifier avec succès (taux d'erreur <4%) des grains de pollens modernes et fossiles, quel que soit leur état de dégradation, ce qui est une avancée majeure. En conclusion, cette thèse contribue à ce que l'analyse des pollens fossiles permette de caractériser les paléoenvironnements avec un degré de précision inégalé à ce jour, et la reconnaissance automatique pourra probablement aider dans un futur proche à reconstituer avec une meilleure résolution spatiale l'habitat des ancêtres de la lignée humaine.

## **Pollen use for the spatial high-resolution reconstruction of Plio-Pleistocene Hominini environments in the East African Rift (Ethiopia and Tanzania)**

This thesis aims at contributing to a better understanding of the micro-habitats associated with the Plio-Pleistocene Hominini of the East African Rift, by means of a study of the pollen signal of modern and fossil sediments in Ethiopia and northern Tanzania. With a modern dataset of 283 surface pollen samples (including 62 new) centred on Ethiopia, an approach has developed that now allows the identification of 24 types of vegetation, including 9 of local riparian vegetation. Applied to the 42 spectra (including 9 new) of the Pliocene Hadar and Busidima Formations (Lower Awash Valley, Ethiopia), this approach shows that between 3.42 and 2.96 Ma, North Afar underwent significant cooling (-6.5 to -10.5°C) associated with the global events of the marine isotopic stages M2, MG2 and MG4. The climatic and vegetation reconstructions induce a succession of very varied environments on a regional scale during this period, while at the same time, locally, the Awash paleoriver and the paleosystems of groundwater sources supported wooded or forest formations at the sites of *Australopithecus afarensis*, thus providing them with a stable wooded habitat. Pollen analysis of a sub-modern sediment core near the shores of Lake Eyasi (Tanzania) sheds light on the vegetation dynamics associated with resurgences and the regional climate in arid zones. Finally, as the reconstructions of paleovegetation associated with the Hominini are broadly limited by the large number of near sterile pollen samples, it was also decided to work on an automated recognition system. Based on convolutional neural networks, the system developed here is capable of successfully identifying (error rate <4%) modern and fossil pollen grains regardless of their state of degradation, which we consider to be a major advance. In conclusion, this thesis contributes to the fact that the analysis of fossil pollen allows the characterization of paleoenvironments with a degree of accuracy difficult to obtain with other proxies to date, and automated recognition will probably help in the near future to reconstitute with a better spatial resolution the habitat of the ancestors of the human lineage.