Letters

Opportunities for unlocking the potential of genomics for African trees

Introduction

Trees (or their absence) represent one of the most defining features of landscapes on the African continent. However, they face major threats including habitat loss and degradation, invasive alien species, disturbance from frequent fire, over-harvesting, pollution, changes in pollinators or dispersers populations, and climate change (Balmford *et al.*, 2001; Davies *et al.*, 2011). Understanding how trees respond to these impacts would require an integrative approach of which genomic science has a potentially major role to play (Plomion *et al.*, 2016).

Since the advent of genomic science, its investigative power has been exploited for trees in temperate regions, particularly involving members of *Pinus*, *Picea*, *Pseudotsuga*, *Populus*, *Eucalyptus*, *Quercus*, *Castanea*, *Malus*, *Prunus*, and *Fraxinus* (Neale & Kremer, 2011; Neale *et al.*, 2013). These species serve as models for exploring various processes in molecular genetics, functional biology, evolutionary biology, phenotypic and genotypic adaptation, physiology and organismal development (Tuskan *et al.*, 2006; Plomion *et al.*, 2016). Although the tropics have exceptionally high tree diversity – with Africa alone having *c.* 50 times more native tree species than temperate Europe (Slik *et al.*, 2015) – tree genomic research in this region lags behind that of temperate ones. Limited funding and the lack of reference genomes for tropical trees have limited the progress of genomic science on the African continent. With over 6000 tree species on the African continent (Slik *et al.*, 2015), there is a need to establish reference genomes for the major tree families.

First, we discuss ways to exploit next generation sequencing (NGS) technologies including genotyping by sequencing (GBS), *de novo* transcriptome assembly and whole genome sequencing to generate genomic resources for nonmodel tree species on the African continent. Second, we discuss landscape genomics, an emerging field in genomic science and discuss research areas in which the genomic resources of trees in Africa can be used to inform research on landscape genomics and to improve food production.

Available DNA resources for African trees

Genomics resources, including reference genomes are completely lacking for African trees. Instead, most available molecular studies of trees in Africa examined a limited number of loci for addressing questions that vary along three major axes: inferring past demographic history, species delimitation/classification, and in conservation (Supporting Information Table S1). In terms of inferring evolutionary history, much of the work addressed the phylogeographic history of dispersal, diversification, vicariance disjunction, evolutionary origin and community phylogenetic structure with focus in central and southern Africa (Table S1). These studies mostly utilized chloroplast DNA regions (usually not more than six) and sometimes nuclear markers, and microsatellites for inferring past demographic history (Table S1). For species delimitation and supraspecific classification, much of the work is grounded on DNA barcoding and disentangling phylogenetic relationships (Table S1). Tree conservation has received the least attention, with existing studies focusing on delineating biodiversity hotspots (Daru *et al.*, 2015a) and conservation genetics (Lowe *et al.*, 2000). Although these studies provided the baseline for further studies, it is clear that the development of DNA resources for African trees is the first step toward unlocking the potential of genomics of trees on the continent.

Development of genomic resources for African trees

Among the common methods for collecting genomic data in nonmodel African tree species without prior molecular resources include amplified fragment length polymorphisms (AFLPs), GBS, diversity arrays technology (DArT), DArTseq, restriction site associated DNA sequencing (RADseq), and *de novo* transcriptome assembly (Fry *et al.*, 2009; Elshire *et al.*, 2011; De Wit *et al.*, 2012; Van Schalkwyk *et al.*, 2012; Li *et al.*, 2015). These methods offer a moderately cheap means of collecting thousands of genomic DNA data points on multiple samples from nonmodel species without reference genomes (Hohenlohe *et al.*, 2010; Elshire *et al.*, 2011; Yeaman *et al.*, 2014).

The GBS technology generates data for thousands of loci per sample by reducing complexity in the genome using restriction enzymes before sequencing (Elshire *et al.*, 2011; Beissinger *et al.*, 2013). A precursor to the GBS was diversity arrays (DArT), a microarray platform for genome wide marker development and analysis. One of the earliest applications of DArT was in a tree species – *Eucalyptus* (Lezar *et al.*, 2004), and it was subsequently widely employed for plant genotyping, often in orphan crops (Gemenet *et al.*, 2015). The GBS version of DArT (DArTseq), provides an order of magnitude more markers (Li *et al.*, 2015). Another variant of the GBS is RADseq, which is being implemented broadly for population genomics of nonmodel species with no prior genomic resources (Andrews *et al.*, 2014). The advantages of these GBS methods are: (1) low cost, (2) not requiring reference genome sequence, (3) treatment of polymorphic tags as dominant markers for many research questions, and (4) can be tailored for each species (Beissinger *et al.*, 2013). The disadvantages include: (1) bioinformatics challenges in alignment of large and polyploid

genomes, (2) bias of genomic regions sampled due to the restriction enzyme–based complexity reduction, and (3) systematic errors for example PCR duplicates introduced during library preparation (Andrews *et al.*, 2014; He *et al.*, 2014).

Another technique to develop genomic resources for African trees is to build a catalog of expressed genes using RNAseq (*de novo* transcriptomes; Mizrachi *et al.*, 2010). This provides information on gene model structure as well as the temporal, spatial and developmental contexts of gene expression. Additionally, we can combine transcriptome with genetic data through expression QTL (quantitative trait locus) analysis to better understand gene regulation, for example in trees growing in different environments (e.g. Munkvold *et al.*, 2013). Genome 'skimming' is another emerging idea that could be applied for conservation of African trees, through the extraction of high-copy sequences, such as plastid genomes, from a low coverage genome (Dodsworth, 2015). This approach may indeed replace the use of PCR-based barcoding techniques (e.g. Gere *et al.*, 2013; Bello *et al.*, 2015), of which the search for a universal barcode for plants is still underway (Hollingsworth *et al.*, 2011).

Finally, if financial resources were available, a good strategy would be to do whole genome sequencing of representative species per family, as carried out for *Eucalyptus grandis* (Myburg *et al.*, 2014). Reference genomes facilitate the speed and accuracy of more cost-effective methods applied to genetically diverse populations, such as GBS, transcriptomics or genome resequencing at low coverage (De Wit *et al.*, 2012).

Future technologies for African tree genomics

Currently, Sanger sequencing remains the gold standard for low throughput sequencing applications such as barcoding (Hollingsworth *et al.*, 2011). However, implementation of short-read NGS technologies such as Illumina, 454 pyrosequencing or Ion Torrent in plant research is growing at an unprecedented rate (Varshney *et al.*, 2014), and holds promise for African trees. The advantages of short-read technologies are high-throughput, cost-effectiveness, low error rate (< 1%; Reuter *et al.*, 2015) compared to other NGS methods, and established analytical tools. For example, a commonly used genome assembler, Velvet, was developed 7 yr ago (Zerbino & Birney, 2008). A disadvantage of short-read technologies is the inability to span regions of repetitive DNA, thus hampering genome assembly (Reuter *et al.*, 2015). Other emerging sequencing technologies that can be applied for African trees include single molecule sequencing available from Pacific Biosciences and Oxford Nanopore which can produce longer reads (usually > 10 kb) (Reuter *et al.*, 2015). A current drawback of single molecule sequencing techniques is the error rate of 5% or more (Reuter *et al.*, 2015). Increased sequencing coverage, for example by sequencing the same molecule multiple times (PacBio), is used to compensate for this, however this makes the cost per base pair greater than short-read NGS. Given that most of the established sequencing technologies are capital intensive, often requiring centralized service providers operating on mega scales, scientists interested in African tree genomics might be faced with the challenge of whether to prioritize technologies or hypothesis-driven

research based on experimental design and data analysis. New innovations include the Oxford Nanopore MinION, which is akin to a 'lab-on-chip' (Watson *et al.*, 2015). It is a cheap and portable (only ~10 cm long) mobile DNA sequencing device powered through a USB port of a laptop, hence ideal for fieldwork.

Prospects for future studies of tree genomics in Africa

The catalogued genomic resources derived from GBS, transcriptomics, and reference whole genome sequencing mentioned earlier can be used for addressing various questions relating to the ecology and evolution of African trees. Here we focus on the prospects of these methods in the study of landscape genomics and to improve food production in Africa.

Landscape genomics of African trees

Landscape genomics is a field that derives from population genetics. It involves the concurrent study of numerous loci from a genome (usually hundreds of markers, genes or genomic regions), both neutral or under adaptive selection for each georeferenced individual across a landscape. Landscape genomics amalgamates the fields of geographic information systems and genomics to explore population dynamics and adaptive genetic variation across landscapes (Schwartz *et al.*, 2010; Bragg *et al.*, 2015). Landscape genomics has been explored in model plants with reference genomes such as *Arabidopsis* (Turner *et al.*, 2010; Fournier-Level *et al.*, 2011; Li *et al.*, 2014), and recently in forest trees (Eckert *et al.*, 2009, 2010; De Kort *et al.*, 2014; Geraldes *et al.*, 2014; McLean *et al.*, 2014). In these studies a strong link between genotype, phenotype and the environment were demonstrated. Furthermore, these studies revealed that genomic data can identify gene regions under positive selection controlling adaptive phenotypic traits in response to environmental variables (see Bragg *et al.*, 2015). The ultimate goal is to precisely quantify population dynamics such as gene flow, genetic differentiation, diversity, and their interaction with several environmental variables in the landscape (Schwartz *et al.*, 2010; Bragg *et al.*, 2015).

Quantifying genomic regions under selective pressures by the environment for landscape genomics involves surveying many loci scattered across the genome of an individual (Manel *et al.*, 2012). Fundamentally, this requires three types of dataset: the genomic data of the species under study, occurrence/abundance data of the species, and environmental variables (e.g. soil type, precipitation, temperature, photoperiod) in the study area (Hand *et al.*, 2015). Climatic data are readily available, for example via WorldClim (Hijmans *et al.*, 2005), and plant distribution data from the Global Biodiversity Information Facility (www.gbif.org). Here, we focus on synthesizing genomic data of African tree species. Since our focus is on landscape genomics of nonmodel African trees without previous genomic resources, the methods chosen should be able to separate neutral from nonneutral markers, and identify suitable environmental variables driving the selection (Foll & Gaggiotti, 2008; Holderegger & Wagner, 2008; Balkenhol *et al.*, 2009). Previous studies have successfully utilized GBS, *de novo* transcriptome assembly, AFLPs, RADseq or single nucleotide

polymorphisms (SNPs) to generate molecular resources for landscape genomics (Schwartz *et al.*, 2010). As a result, the markers of each georeferenced individual studied using any of these methods can be correlated with environmental variables (e.g. substrate, precipitation, temperature, photoperiod, potential evapo-transpiration), to understand the potential adaptive role of genomic regions in response to selective environmental pressure. Here we highlight, as examples, the potential application of landscape genomics for understanding vicariance biogeography, and gregariousness in certain African tree species.

Vicariance biogeography Previous phylogeographic studies of African trees were focused in central Africa (Table S1), but have not been evaluated in a genomic framework. Examples of African trees with potential for exploring phylogeography and landscape genomics include *Vachellia karroo* (= *Acacia karroo*) and *Nymania capensis* (see Fig. 1 for their phylogenetic positions). *Vachellia karroo* (Fabaceae) has a wide distribution (Fig. 2a), traversing several biomes, phytogeographic zones and climatic conditions, including semi-desert (Succulent- and Nama-Karoo), savanna and grassland, as well as both winter and summer-rainfall regimes (White, 1983; Mucina & Rutherford, 2006; Daru *et al.*, 2015b; Fig. 2d). It is a typical so-called chorological and ecological transgressor, several examples which are displayed among African trees (White, 1981). With such wide geographical range, we expect

that the allele frequencies in populations, for example in the winter-rainfall Succulent Karoo, would be different to that of the Highveld grassland with summer-rainfall and winter frost. This is because the Succulent- and Nama-Karoo population would be more adapted to drought tolerance, compared to the Highveld populations that would be adapted to higher rainfall and cold winters, but the genomic basis remains unknown. Study designs for applying landscape genomics to understand phylogeography of nonmodel African trees should use *de novo* transcriptome sequencing to build a catalog library of gene families from various tissues of individuals in the populations of interest. This could have important conservation implications. One can ask, what is the implication of spreading nonnative alleles from a specimen from a population with different environmental tolerance to a new area with another environmental condition, as is often the case with widespread indigenous African species used in horticulture? Another potential tree for exploring phylogeography using landscape genomics is *N. capensis* (Meliaceae) that has two disjunct distributions, one in the Northern Cape of South Africa and southern Namibia and the other in the Western Cape of South Africa (Fig. 2b). Such disjunct distributions between the southeastern and the north-western parts of the subcontinent (a pattern displayed by several plant taxa) was thought to be a result of vicariant events associated with past climatic change (Van Wyk & Smith, 2001). These historically isolated populations may hold
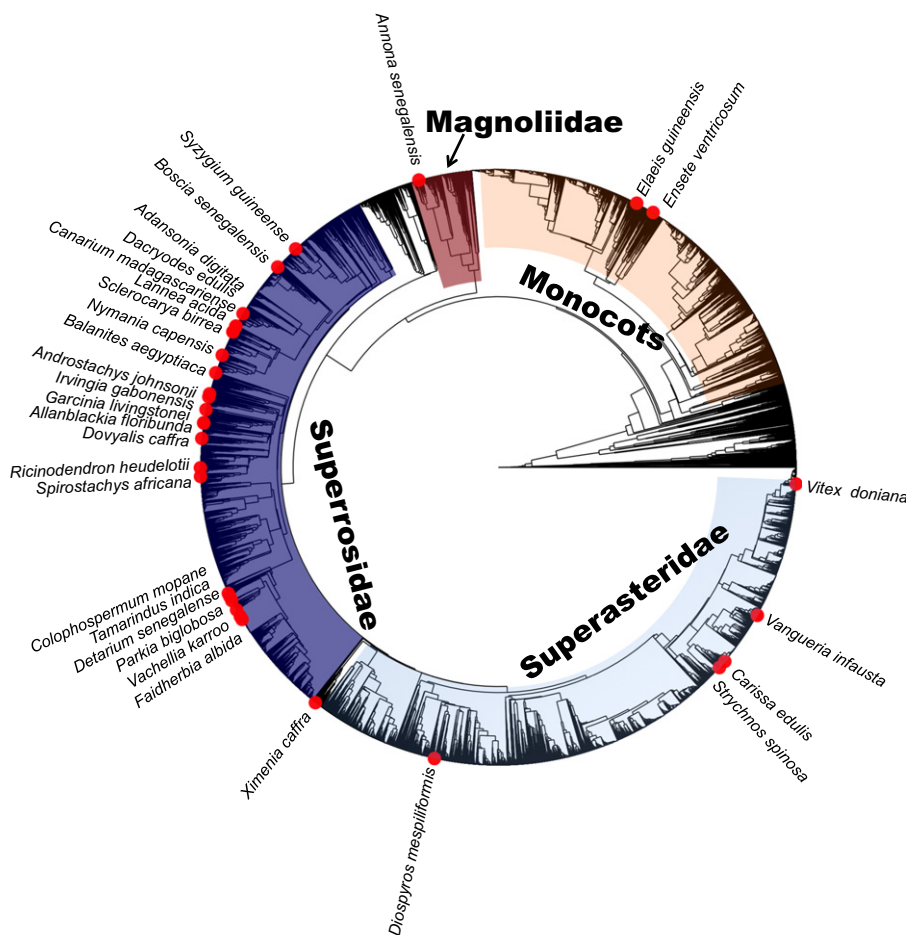

Fig. 1 Dated molecular phylogeny of 31 749 angiosperm species adapted from Zanne *et al.* (2013), with permission. Phylogenetic positions of some of the species with potential for exploring landscape genomics and for improved food production are indicated in red.

information on genetic adaptations resulting from long potential exposure to divergent selection pressures. Within historical biogeography, molecular information may also be useful for weighing the relative merits of hypotheses based on vicariance vs dispersal in explaining geographical distribution of genetic diversity (Avise, 2000). Moreover, molecular information may also assist in the recognition of different taxa within what is currently recognized as single species with disjunct distributions. NGS methods of *de novo* transcriptomes are already being used for species delimitation in temperate regions for nonmodel species, and this could be applied to African trees, with implications for discovering cryptic new infrageneric taxa.

The genomic basis of gregariousness in trees: towards invasion biology The latitudinal gradient of tree diversity has long intrigued biologists (e.g. Slik *et al.*, 2015). The biotic bases for this are not yet fully understood, but genomics as a potential source of information needs to be explored. Many temperate tree species are wind-pollinated (anemophilous) and tend to grow in dense stands, in other words, they are gregarious. For wind pollination to be effective, individuals of a species are expected to occur in close proximity, a trait most probably achieved through the competitive suppression of growth of other species. Hence, based on casual observations of wind-pollinated trees (A. E. van Wyk, unpublished data, 2015), we consider gregariousness to be part of the syndrome for anemophily, a trait usually not considered in the literature on pollination biology (e.g. Friedman & Barrett, 2009). Thus, one would expect plant communities dominated by wind-pollinated trees to be comparatively poor in tree species diversity, as is the case in many temperate regions of the northern hemisphere (Slik *et al.*, 2015). By contrast, trees in Africa are mostly animal-pollinated (Rodger *et al.*, 2004), but with few exceptions of wind pollination, such as *Colophospermum mopane* (Fabaceae), *Androstachys johnsonii* (Picrodendraceae), and *Spirostachys africana* (Euphorbiaceae). *Colophospermum mopane* grows in dense, almost homogeneous stands that cover large parts of southern Africa (Fig. 2c). It is
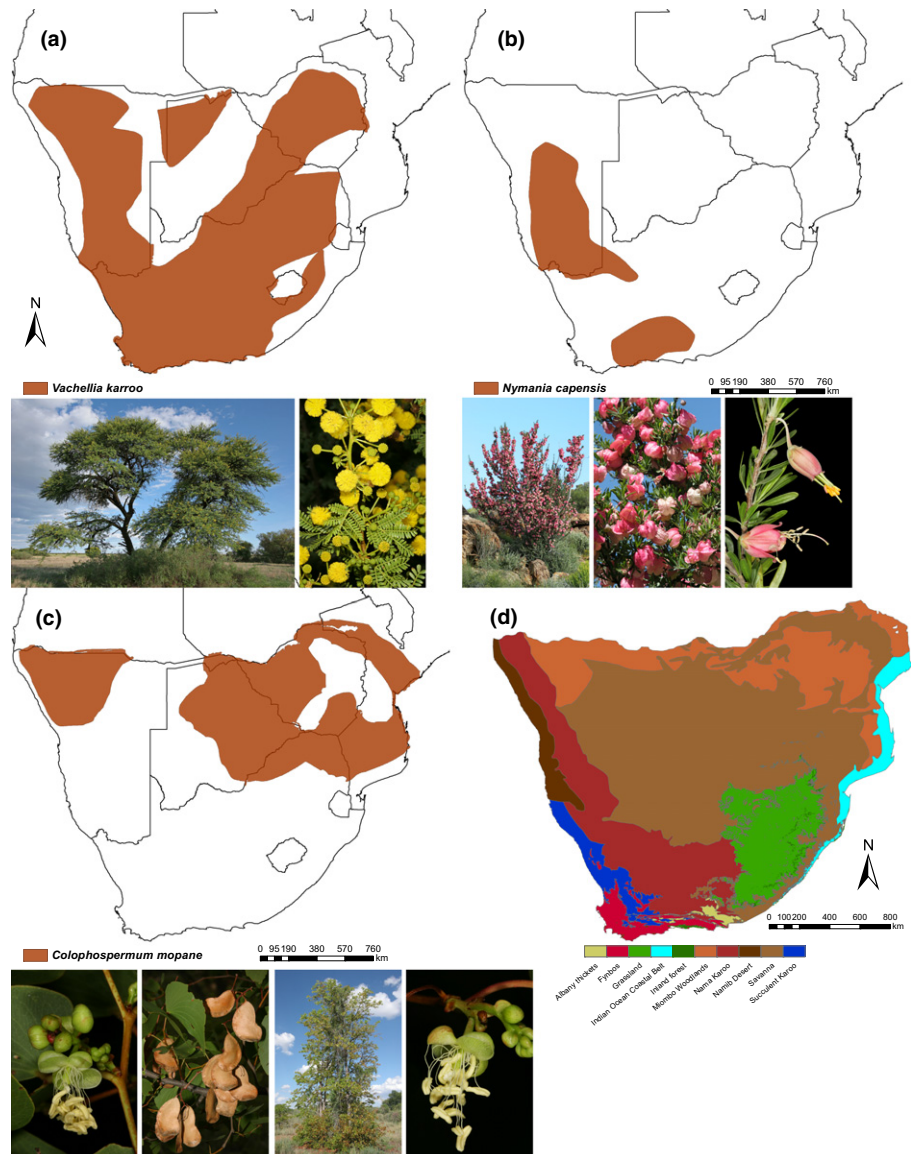
Fig. 2 Examples of potential African tree species for exploring landscape genomics in phylogeography (a, *Vachellia karroo*; b, *Nymania capensis*), a wind-pollinated tree that exhibits gregariousness (c, *Colophospermum mopane*), and currently recognized biomes (d) of southern Africa (from Daru *et al.*, 2015b); the different biomes serve as a reflection of broad-scale environmental heterogeneity in the region. The distributional ranges (shaded) of the trees in southern Africa, as well as their flowers, fruits and habits are depicted.

hypothesized that it exhibits gregariousness by competitively suppressing other woody plants, perhaps by, among others, releasing allelopathic compounds. This could be a good candidate for exploring landscape genomics. For instance, study designs could use AFLPs to establish the different known ecotypes of this species associated with different habitats (Makhado *et al.*, 2014). One could also use *de novo* transcriptomics to build genomic resources towards understanding the genomic basis of gregariousness. This has potentially important implications in ecology, conservation and invasion plant biology (Chown *et al.*, 2015). For example, if the goal is to understand the genomic basis of species invasion, exploring the genomics of gregariousness could lead to the discovery of genes responsible for competitive traits (including allelopathy) during colonization, spread, and suppression of other plants, as is characteristic of many invasive plants (Chown *et al.*, 2015), such as *Lantana camara* and species of *Pinus*. Another implication in this direction is that it could lead to the establishment of a link between pollination syndrome, gregariousness and tree diversity, the existence of which is still hypothetical at this stage.

Improved food production

The exponential population growth rate and the ongoing global climate change is impacting negatively on food production and nutrition in sub-Saharan Africa (Bloom, 2011). While African trees produce various fruits essential for maintaining animal populations (e.g. Breitwisch, 1983; Bleher *et al.*, 2003; Kissling *et al.*, 2008; Daru *et al.*, 2015c), their potential value as food for the increasing human population remains untapped. As a result, genomic science can play a role in improving nutritional quality and increase food production (e.g. see Wang *et al.* (2015) for domestication of maize from teosinte). Thus, genomics can be applied to domestication of African trees with food value in three ways. First, plant breeders can tap into the power of low throughput DNA markers (e.g. PCR based single sequence repeats) or high throughput markers (e.g. SNPs) to select desirable genotypes or counter-select undesirable ones. Second, study designs should use transcriptome sequencing to build a catalog of expressed genes for important African trees, toward discovering genes involved in fruit quality, yield, pest and drought resistance. Third, in cases of barriers to conventional breeding (e.g. requirement for vegetative propagation), the power of genetic modification (GM) can be harnessed via genomics and gene discovery. However, care must be taken to account for ongoing debates on GM crops (Gerasimova, 2015).

The search is in progress for genes under positive selection for desirable traits such as increased crop yield, disease and pest resistance, drought tolerance, and sweetness (Tester & Langridge, 2010; Kanchiswamy *et al.*, 2015). Indeed, on the African continent, one such initiative is the African Orphan Crops Consortium (AOCC; http://africanorphancrops.org/) with the aim of sequencing the genomes of >100 plants commonly used for food in Africa. Specifically, the AOCC aims to develop genomic resources for 'orphan crops' in Africa – crops which are typically neglected and underutilized in regional food security, but have important food or medicinal properties (Naylor *et al.*, 2004). Only 31 species (out of the top 100 plant species) are African trees (Table S2), which is a

promising start. However, these are sparsely distributed across the phylogenetic tree of plant life (Fig. 1), highlighting the need to expand genomics research to cover more representative members of the over 6000 tree species in Africa.

Conclusion

Trees in Africa play keystone roles, having a disproportionately strong influence on co-occurring species, via ecosystem services, worth many millions of dollars; but they are also threatened by environmental and anthropogenic factors. Tapping the potential value of trees and how they respond to these challenges would require an integrative approach of which genomic science has a major role to play. Whereas the availability of reference genomes for many species in the temperate regions has promoted genomics research for temperate trees, here we provided ways to unlock the potential of genomics for nonmodel trees in Africa by focusing on landscape genomics and improved food production and nutrition. While our plea focuses on opportunities for exploiting genomics science for African trees because they represent one of the most defining features of the African landscapes, our suggestions can be extrapolated to other tropical regions such as South America or southeast Asia, or other taxonomic groups including epiphytes, herbs or grasses in the tropics for which genomics resources are lacking. We hope that these suggestions along with the ongoing precipitous drop in DNA sequencing costs and improvements in data quality will open new frontiers in the study of the diversity, ecology and evolution of African trees.

Author contributions

B.H.D. planned and designed the research. B.H.D. analyzed data. B.H.D., D.K.B. and A.E.v.W. contributed materials and analysis tools. B.H.D., D.K.B. and A.E.v.W. wrote the manuscript.

Barnabas H. Daru[1,2]*, Dave K. Berger[1,2,3] and Abraham E. van Wyk[1]

[1]Department of Plant Science, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa;
[2]Genomics Research Institute, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa;
[3]Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa

(*Author for correspondence: tel +27 12 420 2798;
email barnabas.daru@up.ac.za)

References

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G. 2014. Trade-offs and utility of alternative RADseq methods. *Molecular Ecology* 23: 5943–5946.

Avise JC. 2000. *Phylogeography: the history and formation of species.* Cambridge, MA, USA: Harvard University Press.

Balkenhol N, Gugerli F, Cushman S, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH, Participants of the Landscape Genetics Research. 2009. Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24: 455–463.

Balmford A, Moore JL, Brooks T, Burgess N, Hansen LA, Williams P, Rahbek C. 2001. Conservation conflicts across Africa. *Science* 291: 2616–2619.

Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N. 2013. Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* 193: 1073–1081.

Bello A, Daru BH, Stirton CH, Chimphango SBM, van der Bank M, Maurin O, Muthama Muasya AM. 2015. DNA barcodes reveal microevolutionary signals in fire response trait in two legume genera. *AoB Plants* 7: plv124.

Bleher B, Potgieter CJ, Johnson DN, Böhning-Gaese K. 2003. The importance of figs for frugivores in a South African coastal forest. *Journal of Tropical Ecology* 19: 375–386.

Bloom DE. 2011. 7 billion and counting. *Science* 333: 562–569.

Bragg JG, Supple MA, Andrew RL, Borevitz JO. 2015. Genomic variation across landscapes: insights and applications. *New Phytologist* 207: 953–967.

Breitwisch R. 1983. Frugivores at a fruiting *Ficus* vine in a southern Cameroon tropical wet forest. *Biotropica* 15: 125–128.

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA. 2015. Biological invasions, climate change and genomics. *Evolutionary Applications* 8: 23–46.

Daru BH, Van der Bank M, Davies TJ. 2015a. Spatial incongruence among hotspots and complementary areas of tree diversity in southern Africa. *Diversity and Distributions* 21: 769–780.

Daru BH, Van der Bank M, Maurin O, Yessoufou K, Schaefer H, Slingsby JA, Davies TJ. 2015b. A novel phylogenetic regionalization of the phytogeographic zones of southern Africa reveals their hidden evolutionary affinities. *Journal of Biogeography* 43: 155–166.

Daru BH, Yessoufou K, Nuttman C, Abalaka J. 2015c. A preliminary study of bird use of fig *Ficus* species in Amurum Forest Reserve, Nigeria. *Malimbus* 37: 1–15.

Davies TJ, Smith GF, Bellstedt DU, Boatwright JS, Bytebier B, Cowling RM, Forest F, Harmon LJ, Muasya AM, Schrire BD et al. 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biology* 9: e1000620.

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology* 23: 4709–4721.

De Wit P, Pespeni MH, Ladner JT, Barshis DJ, Seneca F, Jaris H, Therkildsen NO, Morikawa M, Palumbi SR. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12: 1058–1067.

Dodsworth S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science* 20: 525–527.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St Clair JB, Neale DB. 2009. Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* 182: 1289–1302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB. 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185: 969–982.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86–89.

Friedman J, Barrett SCH. 2009. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Annals of Botany* 103: 1515–1527.

Fry NK, Savelkoul PH, Visca P. 2009. Amplified fragment-length polymorphism analysis. *Methods in Molecular Biology* 551: 89–104.

Gemenet DC, Leiser WL, Zangre RG, Angarawai II, Sanogo MD, Sy O, Mahendar T, Hash CT, Haussmann BIG. 2015. Association analysis of low-phosphorus tolerance in West African pearl millet using DArT markers. *Molecular Breeding* 35: 1–20.

Geraldes A, Farzaneh N, Grassa CJ, McKown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* 68: 3260–3280.

Gerasimova K. 2015. Debates on genetically modified crops in the context of sustainable development. *Science and Engineering Ethics*. doi: 10.1007/s11948-015-9656-y.

Gere J, Yessoufou K, Daru BH, Maurin O, Mankga LT, van der Bank M. 2013. Incorporating *trnH-psbA* to core DNA barcodes improves discrimination of species within southern African Combretaceae. *ZooKeys* 365: 127–147.

Hand BK, Lowe WH, Kovach RP, Muhlfeld CC, Luikart G. 2015. Landscape community genomics: understanding eco-evolutionary processes in complex environments. *Trends in Ecology & Evolution* 30: 161–168.

He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5: 484.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6: e1000862.

Holderegger R, Wagner HE. 2008. Landscape genetics. *BioScience* 58: 199–207.

Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and using a plant DNA barcode. *PLoS ONE* 6: e19254.

Kanchiswamy CN, Sargent DJ, Velasco R, Maffei ME, Malnoy M. 2015. Looking forward to genetically edited fruit crops. *Trends in Biotechnology* 33: 62–64.

Kissling WD, Field R, Böhning-Gaese K. 2008. Spatial patterns of woody plant and bird diversity: functional relationships or environmental effects? *Global Ecology and Biogeography* 17: 327–339.

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD. 2004. Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. *Theoretical and Applied Genetics* 109: 1329–1336.

Li H, Vikram P, Singh RP, Kilian A, Carling J, Song J, Burgueno-Ferreira JA, Bhavani S, Huerta-Espino J, Payne T et al. 2015. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics* 16: 1–15.

Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. 2014. Genetic variation for life history sensitivity to seasonal warming in *Arabidopsis thaliana*. *Genetics* 196: 569–577.

Lowe AJ, Gillies ACM, Wilson J, Dawson IK. 2000. Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. *Molecular Ecology* 9: 831–841.

Makhado RA, Mapaure I, Potgieter MJ, Luus-Powell WJ, Saidi AT. 2014. Factors influencing the adaptation and distribution of *Colophospermum mopane* in southern Africa's mopane savannas – a review. *Bothalia* 44: art. 152.

Manel S, Albert CH, Yoccoz NG. 2012. Sampling in landscape genomics. In: Pompanon F, Bonin A, eds. *Data production and analysis in population genomics: methods and protocols, methods in molecular biology, vol. 888.* New York, NY, USA: Springer Science+Business Media, 93–112.

McLean EH, Prober SM, Stock WD, Steane DA, Potts BM, Vaillancourt RE, Byrne M. 2014. Plasticity of functional traits varies clinally along a rainfall gradient in *Eucalyptus tricarpa*. *Plant, Cell & Environment* **37**: 1440–1451.

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. 2010. *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics* **11**: 681.

Mucina L, Rutherford MC. 2006. *The vegetation of South Africa, Lesotho and Swaziland*. Pretoria, South Africa: South African National Biodiversity Institute.

Munkvold JD, Laudencia-Chingcuanco D, Sorrells ME. 2013. Systems genetics of environmental response in the mature wheat embryo. *Genetics* **194**: 265–277.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.

Naylor RL, Falcon WP, Goodman RM, Jahn MM, Sengooba T, Tefera H, Nelson RJ. 2004. Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* **29**: 15–44.

Neale D, Langley C, Salzberg S, Wegrzyn J. 2013. Open access to tree genomes: the path to a better forest. *Genome Biology* **14**: 120.

Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**: 111–122.

Plomion C, Bastien C, Bogeat-Triboulot MB, Bouffier L, Déjardin A, Duplessis S, Fady B, Heuertz M, Le Gac AL, Le Provost G *et al.* 2016. Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* **73**: 77–103.

Reuter JA, Spacek DV, Snyder MP. 2015. High-throughput sequencing technologies. *Molecular Cell* **58**: 586–597.

Rodger JG, Balkwill K, Gemmill B. 2004. African pollination studies: where are the gaps? *International Journal of Tropical Insect Science* **24**: 5–28.

Schwartz MK, McKelvey KS, Cushman SA, Luikart G. 2010. Landscape genomics: a brief perspective. In: Cushman SA, Huettmann F, eds. *Spatial complexity, informatics, and wildlife conservation*. Tokyo, Japan: Springer, 165–174.

Slik JW, Arroyo-Rodríguez V, Aiba S, Alvarez-Loayza P, Alves LF, Ashton P, Balvanera P, Bastian ML, Bellingham PJ, van den Berg E *et al.* 2015. An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences, USA* **112**: 7472–7477.

Tester M, Langridge P. 2010. Breeding technologies to increase crop production in a changing world. *Science* **327**: 818–822.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260–263.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). *Science* **313**: 1596–1604.

Van Schalkwyk A, Wenzl P, Smit S, Lopez-Cobollo R, Kilian A, Bishop G, Hefer C, Berger D. 2012. Bin mapping of tomato diversity array (DArT) markers to genomic regions of *Solanum lycopersicum × Solanum pennellii* introgression lines. *Theoretical and Applied Genetics* **124**: 947–956.

Van Wyk AE, Smith G. 2001. *Regions of floristic endemism in southern Africa. A review with emphasis on succulents*. Hatfield, South Africa: Umdaus Press.

Varshney RK, Terauchi R, McCouch SR. 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biology* **12**: e1001883.

Wang H, Studer AJ, Zhao Q, Meeley R, Doebley JF. 2015. Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in *tga1*. *Genetics* **200**: 965–974.

Watson M, Thomson M, Risse J, Talbot R, Santoyo-Lopez J, Gharbi K, Blaxter M. 2015. poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics* **31**: 114–115.

White F. 1981. The history of the Afromontane archipelago and the scientific need for its conservation. *African Journal of Ecology* **19**: 33–54.

White F. 1983. *The vegetation of Africa. A descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa*. Paris, France: UNESCO.

Yeaman S, Hodgins KA, Suren H, Nurkowski KA, Rieseberg LH, Holliday JA, Aitken SN. 2014. Conservation and divergence of gene expression plasticity following *c.* 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca × Picea engelmannii*). *New Phytologist* **203**: 578–591.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2013. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**: 89–92.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**: 821–829.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Summary of findings of some previous molecular studies of trees/woody species across Africa in the fields of phylogeography, molecular systematics and phylogenetics (e.g. DNA barcoding)

Table S2 List of African trees with food value prioritized for genomics analysis by the African Orphan Crops Consortium (http://africanorphancrops.org/)

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: anemophily, food plants, gregariousness, landscape genomics, orphan crops, transcriptomes, trees of Africa, vicariance.