

Les paysages de recombinaison de la Cione

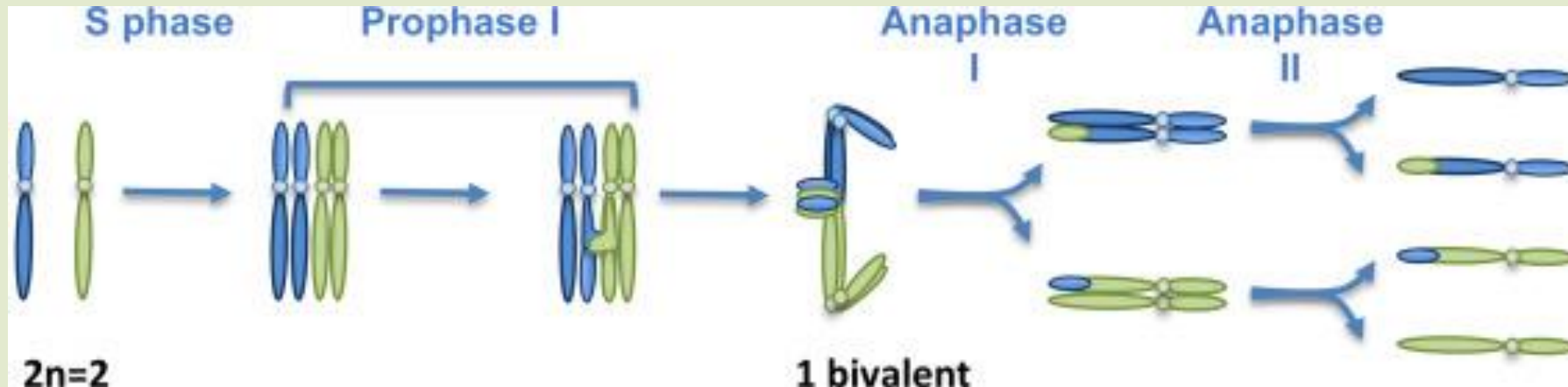
Stage de M1 – Parcours BiM de l'INSA Lyon

Aurélie Fischer, encadrée par Laurent Duret et Julien Joseph

1 - Introduction

Pourquoi s'intéresser aux paysages de recombinaison ?

- La recombinaison meiotique

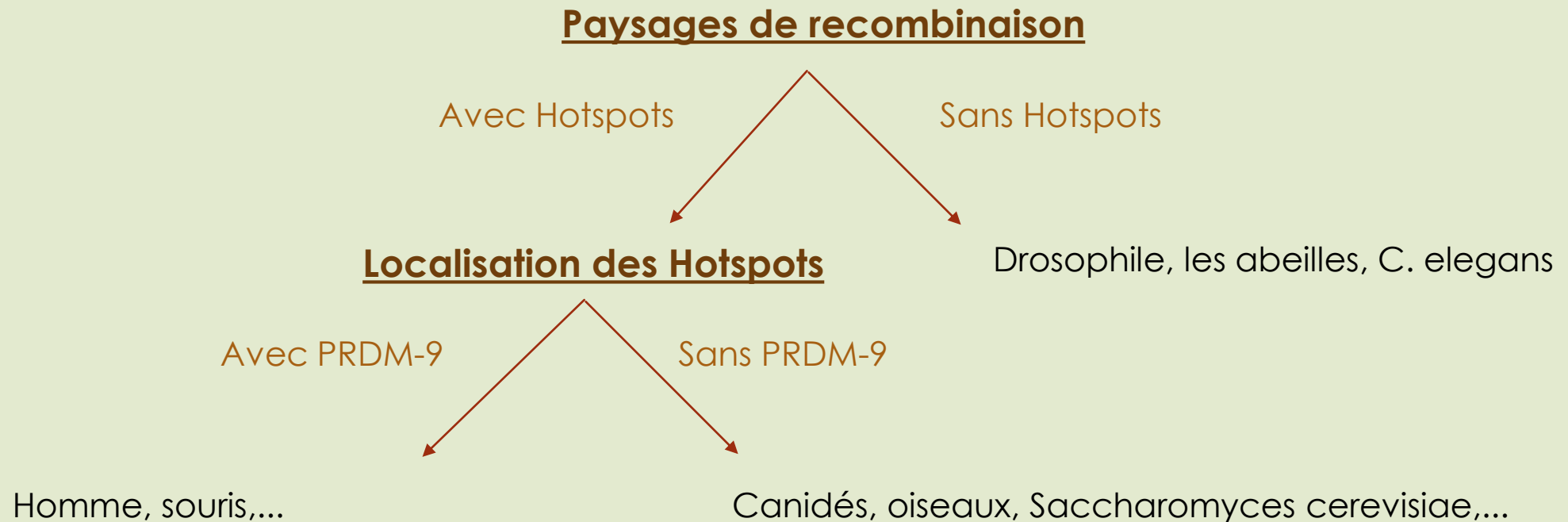


Mathilde Grelon, *Meiotic recombination mechanisms*, Comptes Rendus Biologies, Volume 339, Issues 7–8, 2016, Pages 247-251, ISSN 1631-0691, <https://doi.org/10.1016/j.crv.2016.04.003>.

1 - Introduction

Pourquoi s'intéresser aux paysages de recombinaison ?

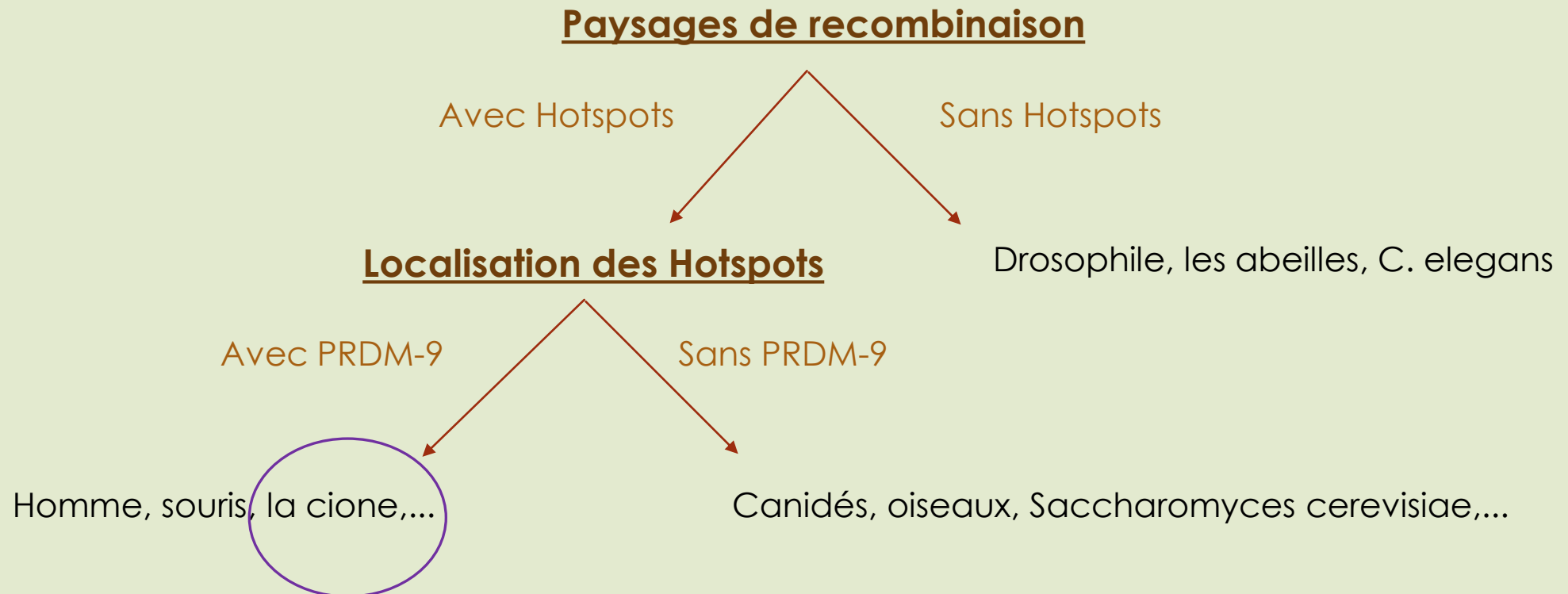
- Comprendre les mécanismes de recombinaison :



1 - Introduction

Pourquoi s'intéresser aux cartes de recombinaison ?

- Comprendre les mécanismes de recombinaison :



1 - Introduction

Comment réaliser une carte de recombinaison ?

- Méthodes par pédigré
 - S'appuient sur des expériences avec des croisements
 - **Faible résolution** de carte, mais plus fiables
- Méthodes plus récentes
 - Algorithmes qui se basent sur les données de polymorphisme
 - Calcul du déséquilibre de liaison
 - **Haute résolution** de carte

1 - Introduction

Comment réaliser une carte de recombinaison ?

- Méthodes par pédigré
 - S'appuient sur des expériences avec des croisements
 - **Faible résolution** de carte, mais plus fiables
- Méthodes plus récentes
 - Algorithmes qui se basent sur les données de polymorphisme
 - Calcul du déséquilibre de liaison
 - **Haute résolution** de carte

Problématique :

Est-ce que les méthodes basées sur le déséquilibre de liaison permettent de construire une carte de recombinaison fiable de la Cione ?

2 - Matériel et méthodes

Données de départ : **la Cione**

- Génome de 100Mb avec 14 Chromosomes
- 2 espèces à analyser :
 - 13 individus de *Ciona intestinalis*
 - 11 individus de *Ciona robusta*



C.intestinalis - © Y. Fontana / Station biologique de Roscoff (CNRS/UPMC)

On utilise les données de polymorphisme pour faire les cartes

- Etapes déjà réalisées : Séquençage → Alignement → Détection de variants
- Les variants répertoriés sont les SNPs (Single Nucleotide Polymorphism)

2 - Matériel et méthodes



2 - Matériel et méthodes

Etape 1 – Filtrage des SNPs

On ne garde que les SNPs qui sont :

- Bialléliques
- Sans insertions/délétions (indels)
- Fréquence allélique $> 10 \%$
- Avec au maximum 50% de génotype manquant

2 - Matériel et méthodes

4 filtres :

- Bialléliques
- Sans indels
- Fréquence Allélique > 0.1
- 50 % max de génotype manquant

-1-
DONNEES DE
POLYMORPHISME

Filtrage des
SNPs

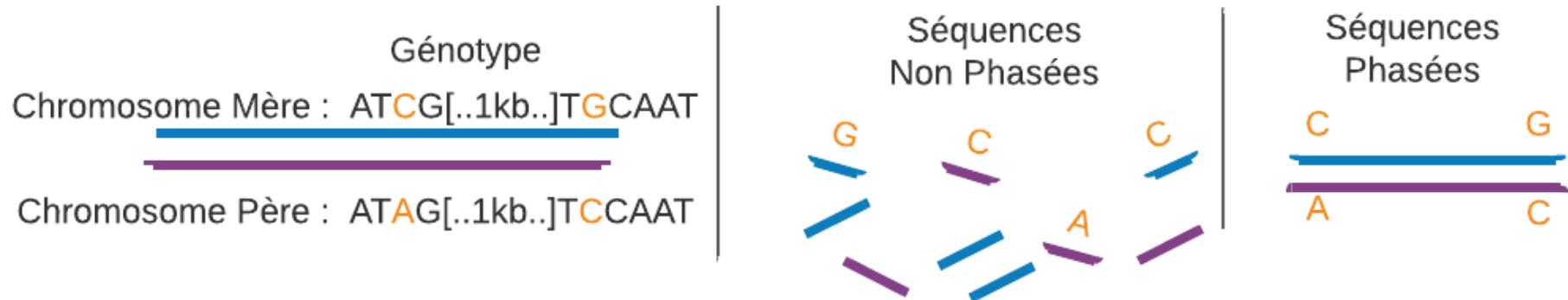
Phasage des
SNPs

Inférence des
taux de
recombinaison

2 - Matériel et méthodes

Etape 2 – Phasage des SNPs

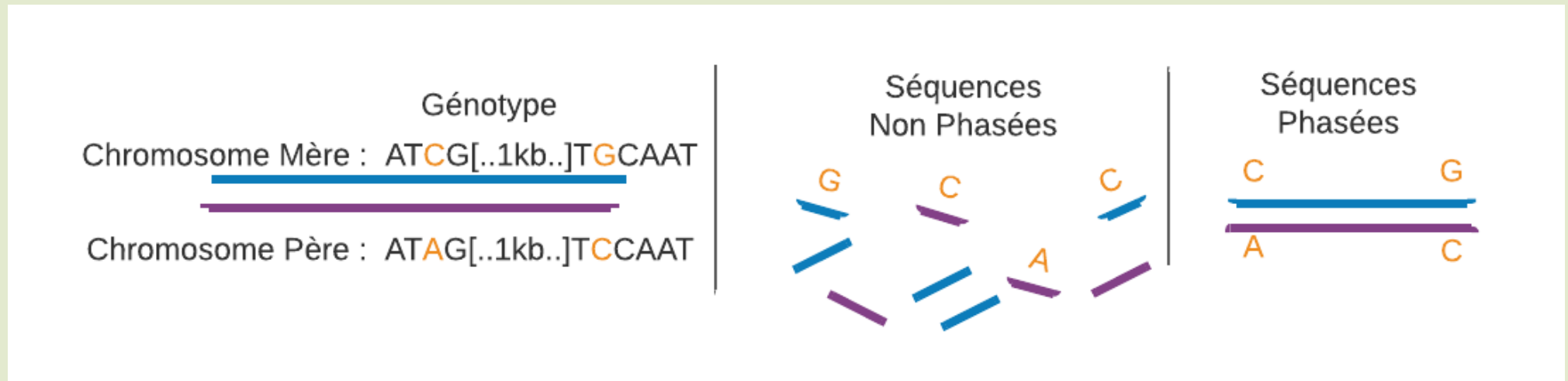
Qu'est-ce que le phasage ?



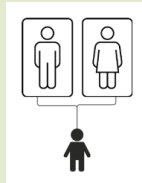
2 - Matériel et méthodes

Etape 2 – Phasage des SNPs

Qu'est-ce que le phasage ?



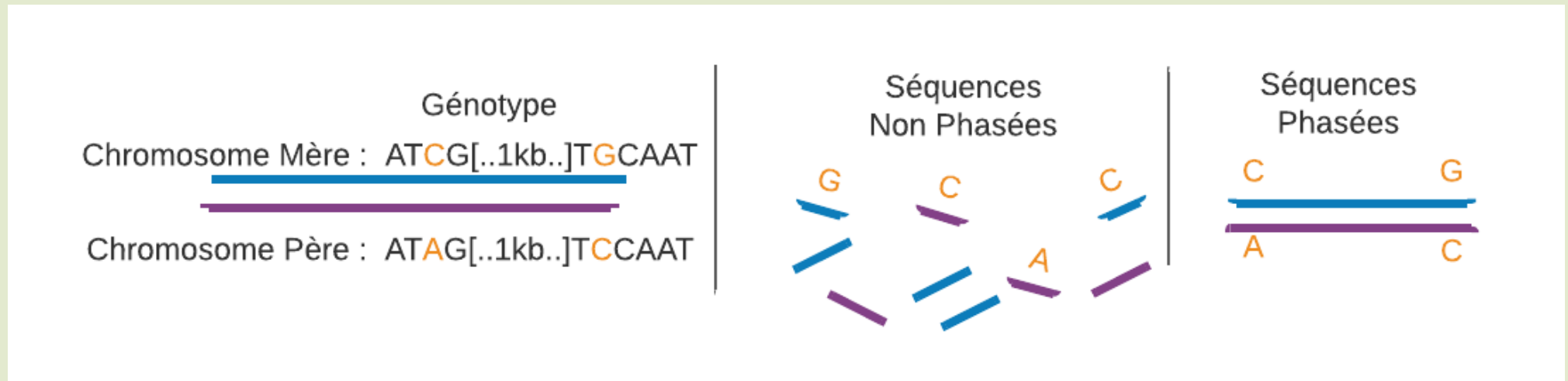
Phasage par trio



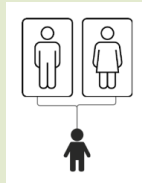
2 - Matériel et méthodes

Etape 2 – Phasage des SNPs

Qu'est-ce que le phasage ?



Phasage par trio



Phasage statistique

2 - Matériel et méthodes

4 filtres :

- Bialléliques
- Sans indels
- Fréquence Allélique > 0.1
- 50 % max de génotype manquant

-1-
DONNEES DE
POLYMORPHISME

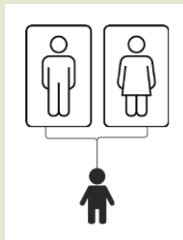
Filtrage des
SNPs

Phasage des
SNPs

Inférence des
taux de
recombinaison

2 types de phasages

Phasage par trio



Phasage statistique

≈

2 - Matériel et méthodes

Etape 3 – Inférence des taux de recombinaison

On utilise 2 outils pour construire les cartes de recombinaison :

- LDhat
- LDhelmet

2 - Matériel et méthodes

Etape 3 – Inférence des taux de recombinaison

On utilise 2 outils pour construire les cartes de recombinaison :

- LDhat
- LDhelmet

Remarque : les 2 outils emploient des algorithmes rjMCMC qui introduisent de la stochasticité dans les résultats.

→ Les cartes ne sont pas parfaitement reproductibles à partir d'un même jeu de données en entrée.

2 - Matériel et méthodes

4 filtres :

- Bialléliques
- Sans indels
- Fréquence Allélique > 0.1
- 50 % max de génotype manquant

2 outils :

- LDhat
- LDhelmet

→ **Algos rjMCMC**

-1-
DONNEES DE
POLYMORPHISME

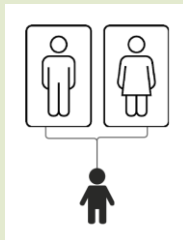
Filtrage des
SNPs

Phasage des
SNPs

Inférence des
taux de
recombinaison

2 types de phasages

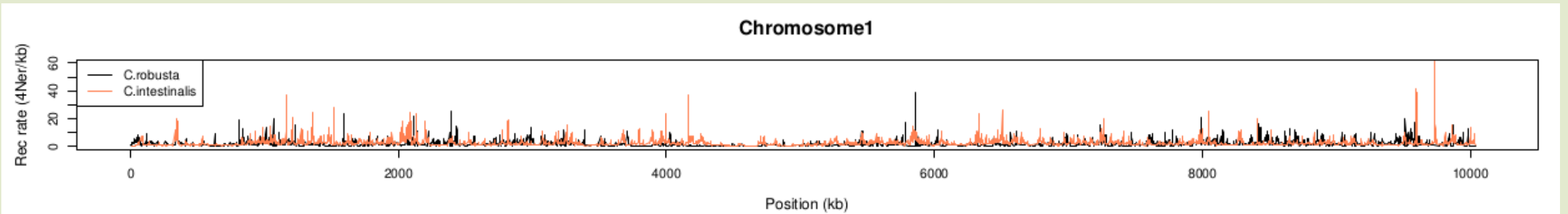
Phasage par trio



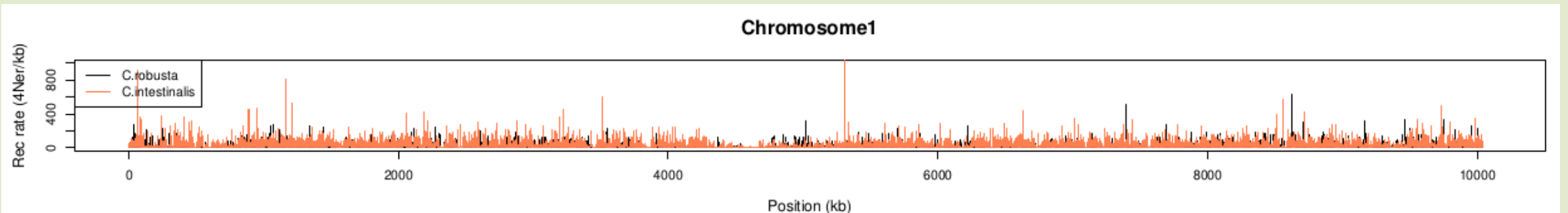
Phasage statistique

≈

3 - Résultats



Carte de recombinaison n°1 - SNPs phasés par trio – Taux inférés par LDhat



Carte de recombinaison n°2 - SNPs phasés statistiquement – Taux inférés par LDhat

3 - Résultats

Différents aspects à regarder :

- Reproductibilité des cartes
- Comparaison entre le phasage statistique et le phasage par trio
- Comparaison entre LDhat et LDhelmet

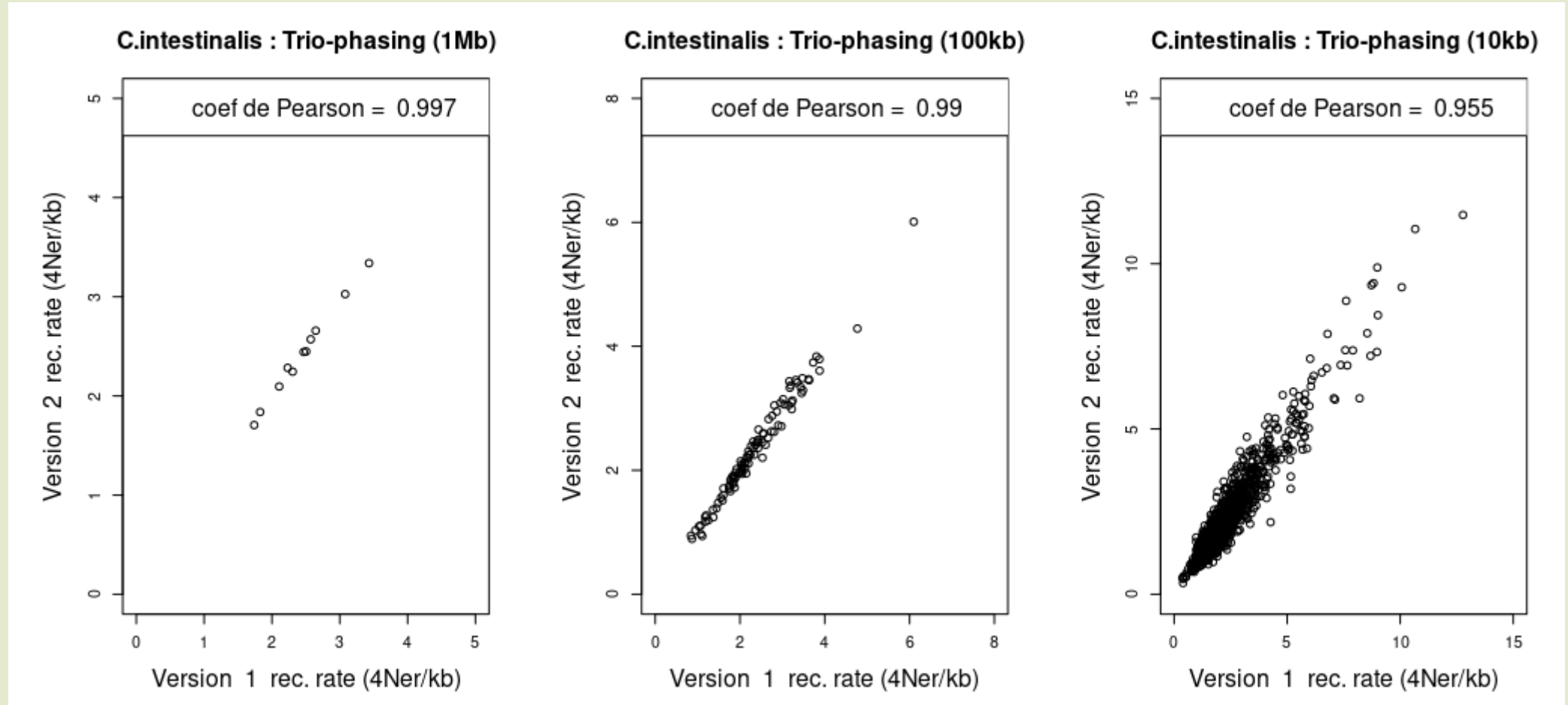
3 - Résultats

Différents aspects à regarder :

- **Reproductibilité des cartes**
- Comparaison entre le phasage statistique et le phasage par trio
- Comparaison entre LDhat et LDhelmet

3 - Résultats

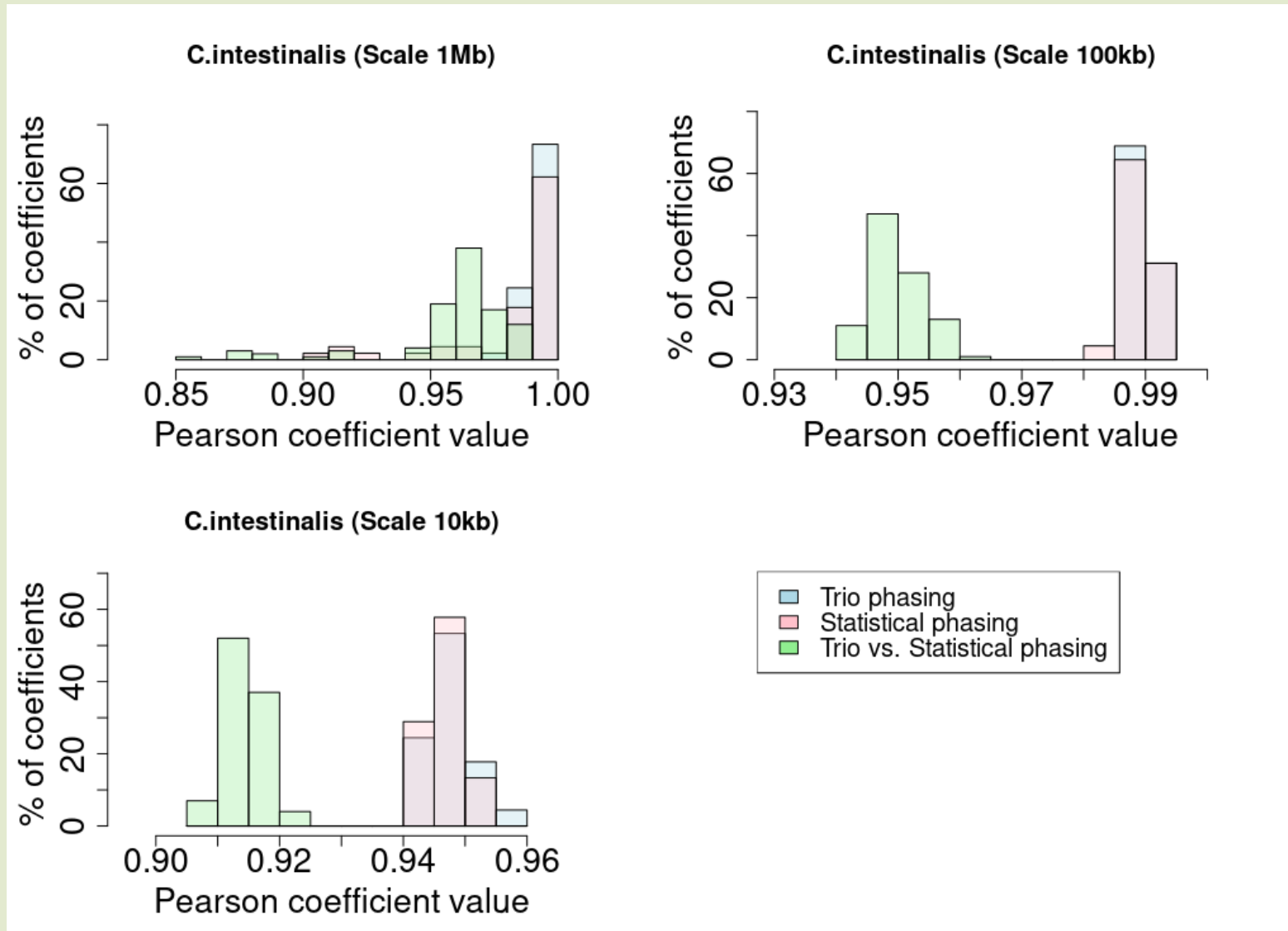
Reproductibilité des cartes chez *C.intestinalis*



Corrélation entre 2 réplicats de carte
du Chromosome 1 de *C. intestinalis*, avec les SNPs phasés par trio, sous LDhat

3 - Résultats

Reproductibilité des cartes chez *C.intestinalis*



Distribution des coefficients de Pearson

Corrélation entre 10 réplicats de carte de recombinaison du chromosome 1 de *Ciona intestinalis* réalisés sous LDhat, avec 2 jeux de SNPs

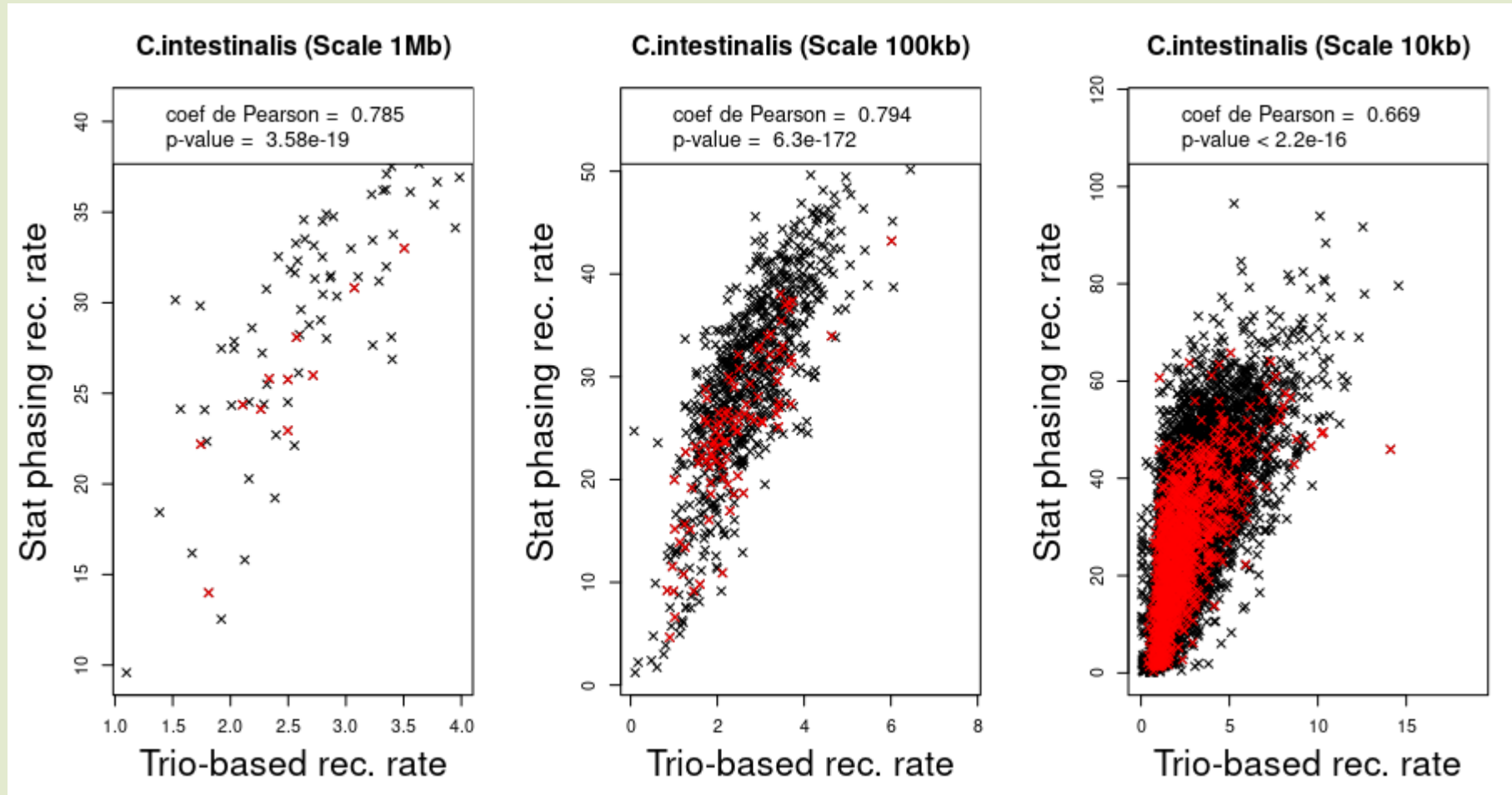
3 - Résultats

Différents aspects à regarder :

- Reproductibilité des cartes
- **Comparaison entre le phasage statistique et le phasage par trio**
- Comparaison entre LDhat et LDhelmet

3 - Résultats

Phasage par trio vs. phasage statistique chez *C.intestinalis*



Corrélation entre les deux cartes de recombinaison de *C.intestinalis* faites avec Ldhat (taux en rho/kb), l'une avec phasage statistique, l'autre avec phasage par trio.

Les points en rouge correspondent au chromosome 1.

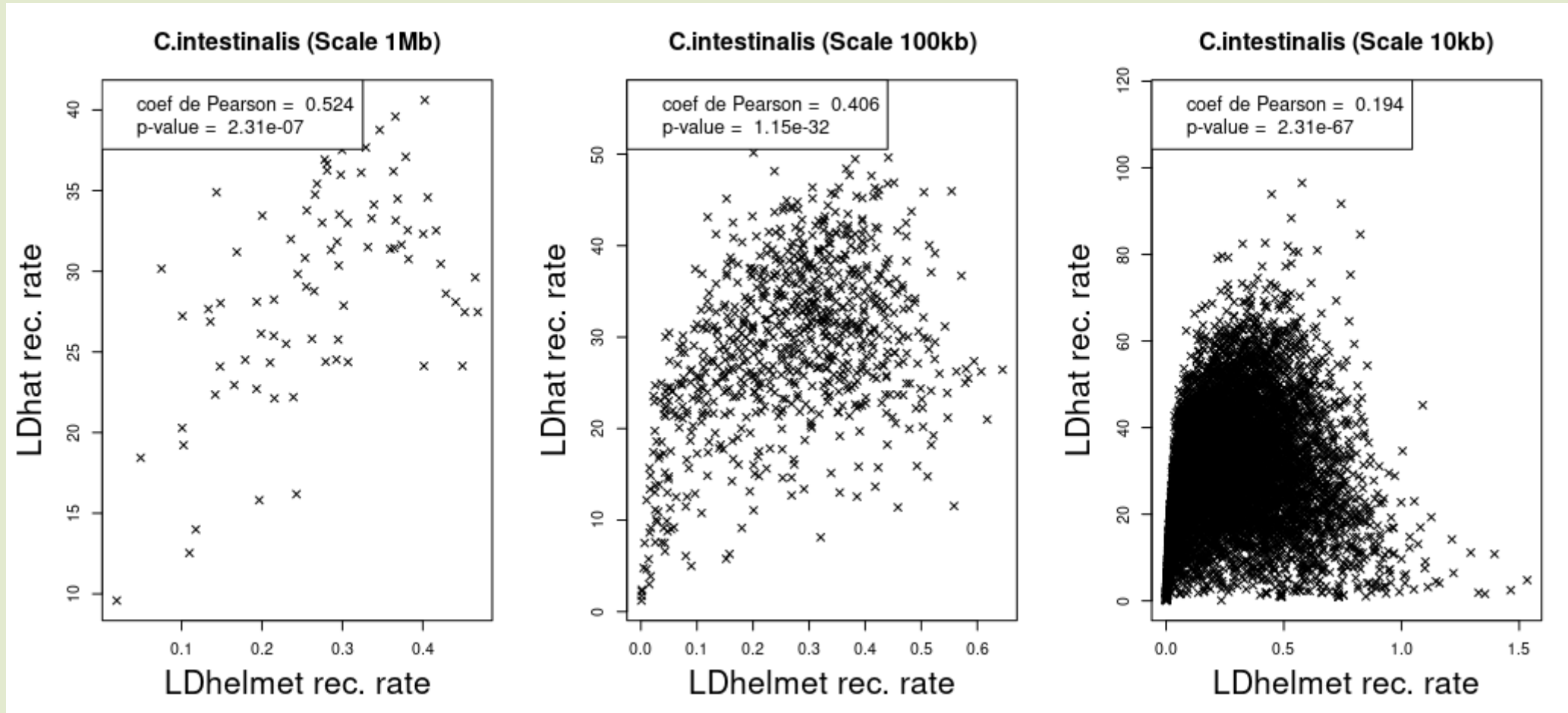
3 - Résultats

Différents aspects à regarder :

- Reproductibilité des cartes
- Comparaison entre le phasage statistique et le phasage par trio
- **Comparaison entre LDhat et LDhelmet**

3 - Résultats

LDhat vs. LDhelmet chez *C.intestinalis*



Corrélation entre les deux cartes de recombinaison de *C.intestinalis* faites avec phasage statistique, l'une avec Ldhat (taux en rho/kb), l'autre avec Ldhelmet (taux en rho/bp)

4 - Conclusion

But : Est-ce que les méthodes basées sur le déséquilibre de liaison permettent de construire une carte de recombinaison fiable de la Cione ?

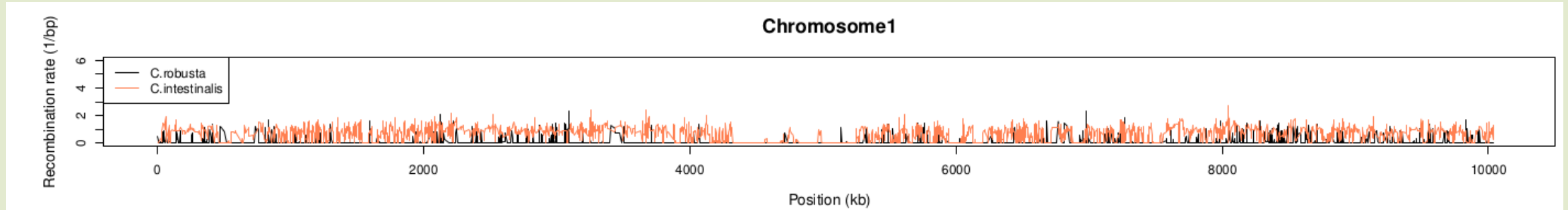
Conclusions :

- Les cartes sont quasiment reproductibles malgré la stochasticité introduite par les algorithmes rjMCMC
- Le phasage statistique : un bon compromis pour gagner en résolution sur la carte
- LDhat et LDhelmet donnent des résultats très différents → au moins une des cartes est fausse.

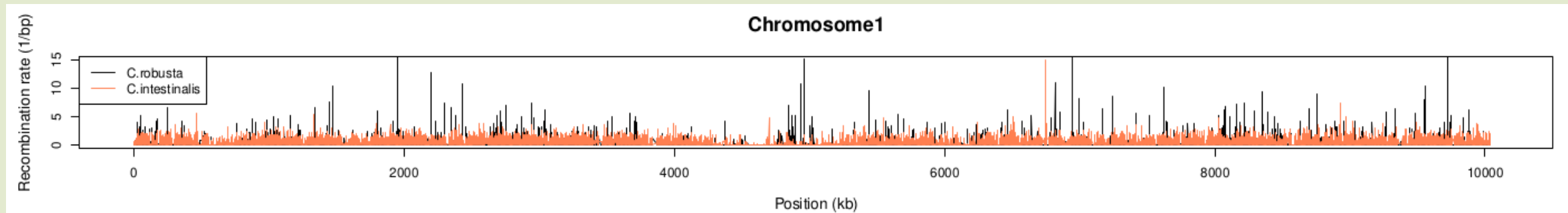
**Merci de
votre
attention**

Annexes

Annexe 1- Résultats LDhelmet



Carte de recombinaison n°3 - SNPs phasés par trio – Taux inférés par LDhelmet

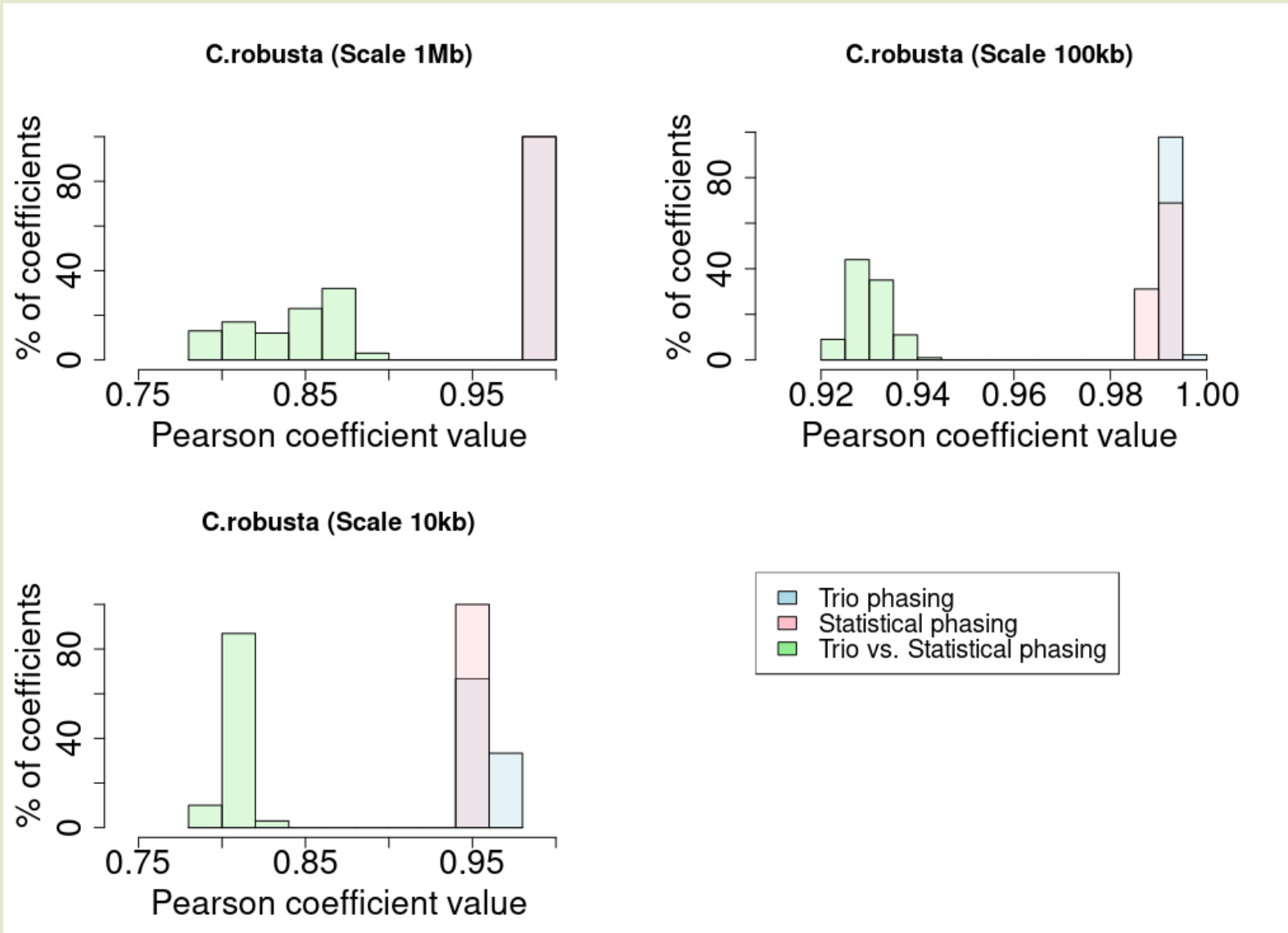


Carte de recombinaison n°4 - SNPs phasés statistiquement – Taux inférés par LDhelmet

Annexe 2

31

Reproductibilité des cartes chez *C.robusta*

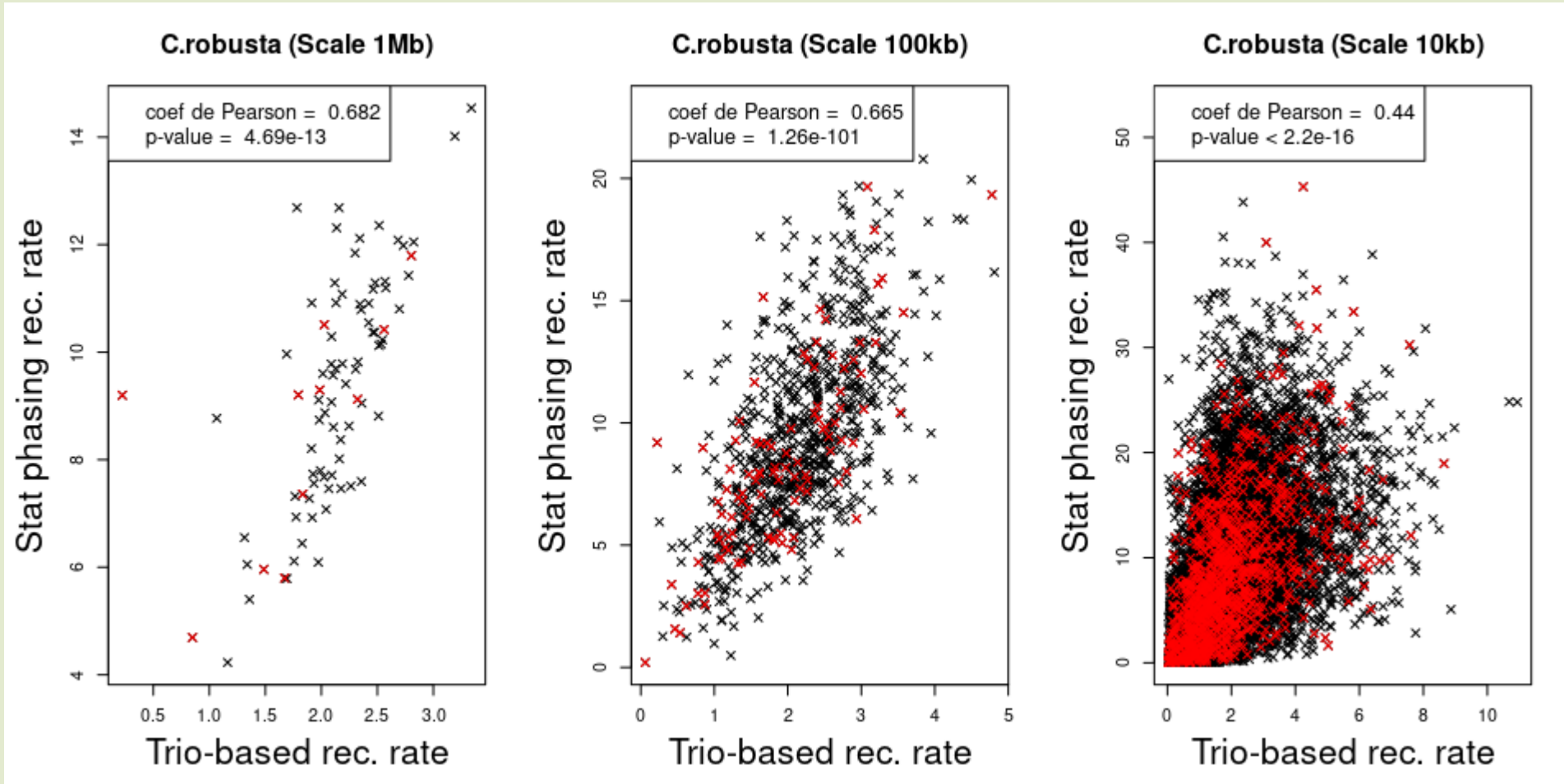


Distribution des coefficients de Pearson

Corrélation entre 10 réplicats de carte de recombinaison du chromosome 1 de *Ciona robusta* réalisés sous LDhat, avec 2 jeux de SNPs

Annexe 3

Phasage par trio vs. phasage statistique chez *C.robusta*

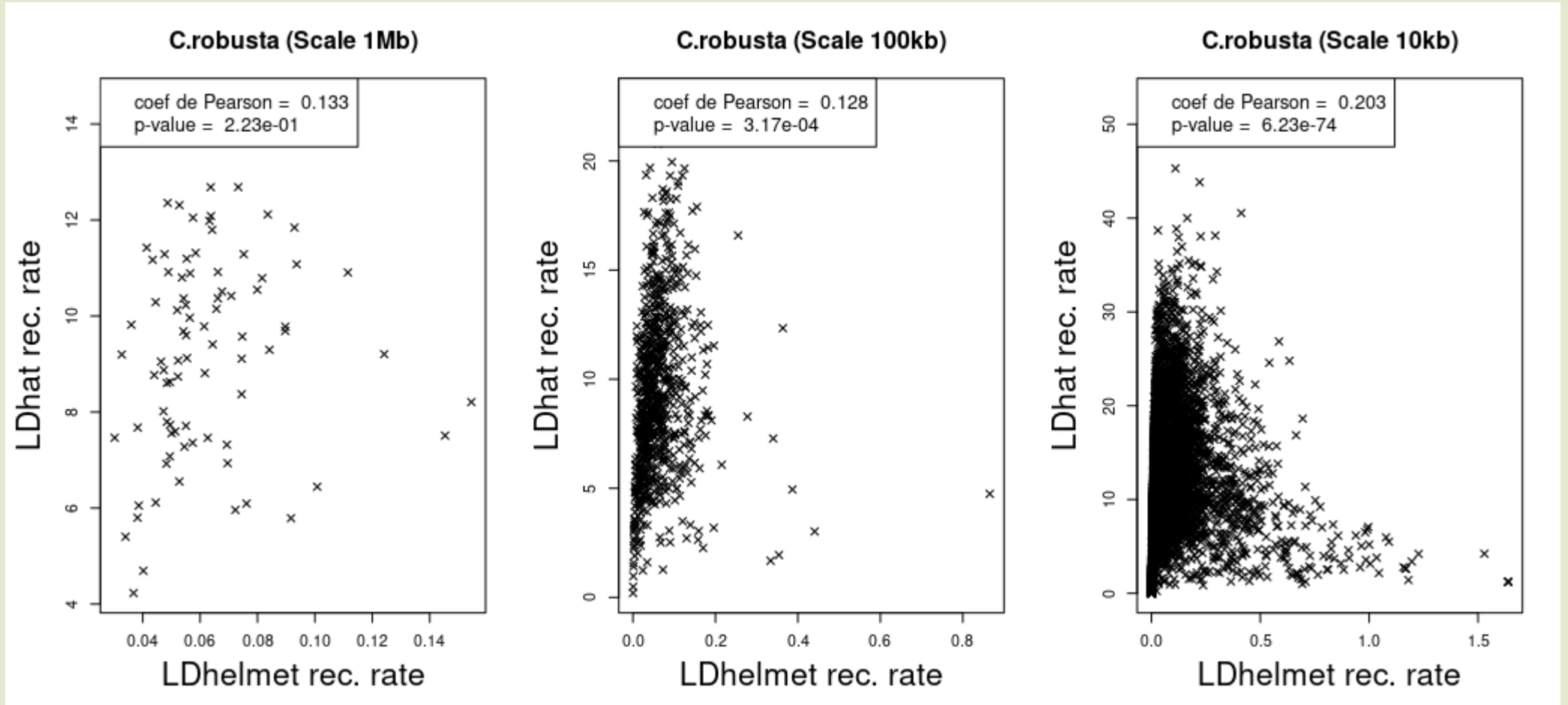


Corrélation entre les deux cartes de recombinaison de *C.robusta* faites avec LDhat, l'une avec phasage statistique, l'autre avec phasage par trio.

Les points en rouge correspondent au chromosome 1.

Annexe 4

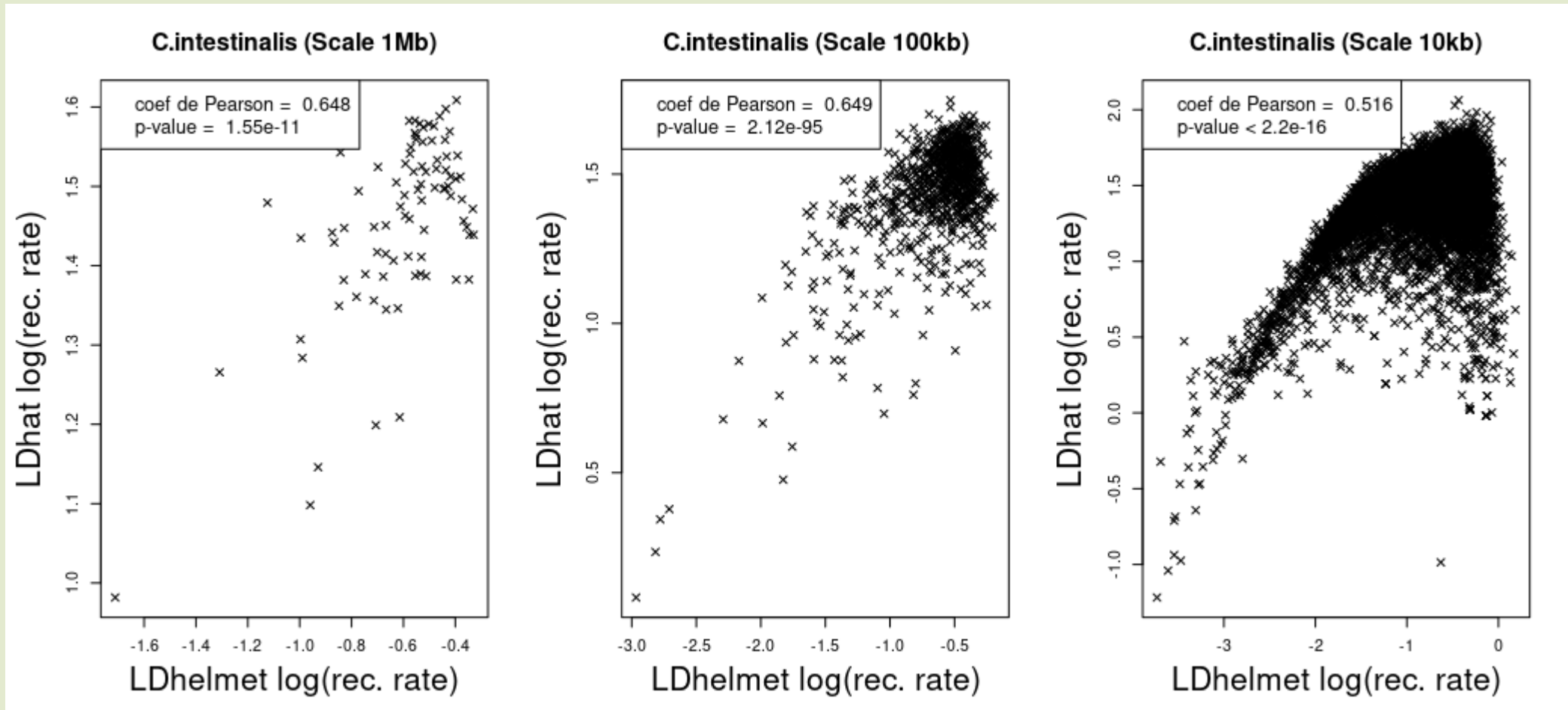
LDhat vs. LDhelmet chez *C.robusta*



Corrélation entre les deux cartes de recombinaison de *C.robusta* faites avec phasage statistique, l'une avec LDhat, l'autre avec LDhelmet

Annexe 5

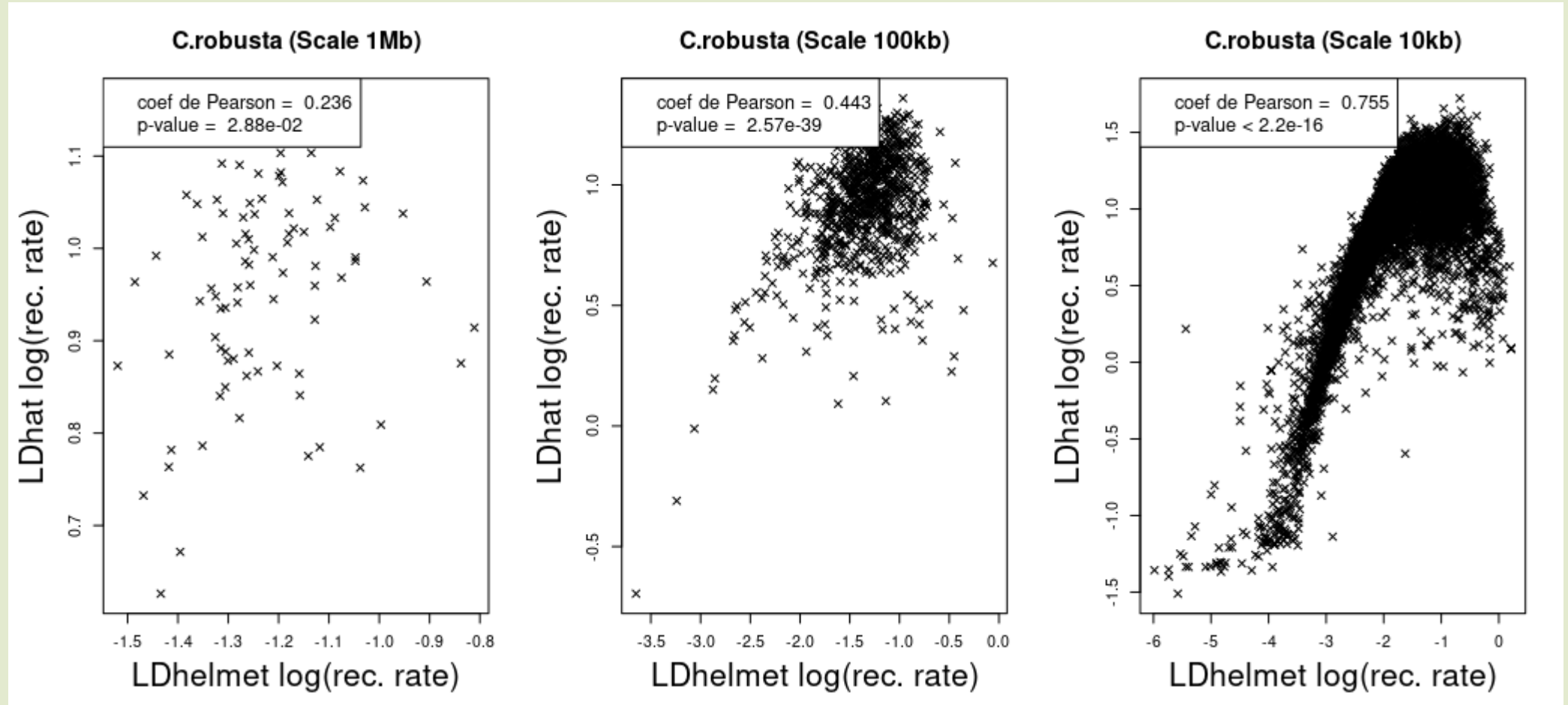
LDhat vs. LDhelmet chez *C.intestinalis* (échelle log)



Corrélation entre les deux cartes de recombinaison de *C.intestinalis* faites avec phasage statistique, l'une avec LDhat, l'autre avec LDhelmet (échelle logarithmique)

Annexe 6

LDhat vs. LDhelmet chez *C.robusta* (échelle log)



Corrélation entre les deux cartes de recombinaison de *C.robusta* faites avec phasage statistique, l'une avec LDhat, l'autre avec LDhelmet (échelle logarithmique)

4 – Discussion

36

Quelle carte est la plus fiable ?

On se base sur les critères suivants :

- Corrélation avec le taux de GC
- Corrélation avec la longueur des chromosomes

Annexe 7A

37

Quelle carte est la plus fiable ? - Critère du taux de GC

	Echelle	1Mb	100kb	10kb
<u>Carte n°1</u> Phasage par trio LDhat	<i>C.intestinalis</i>	0.19	0.31*	0.25*
	<i>C.robusta</i>	0.02	-0.04	-0.09
<u>Carte n°2</u> Phasage statistique LDhat	<i>C.intestinalis</i>	0.37*	0.30*	0.22*
	<i>C.robusta</i>	0.10	-0.13*	-0.19*
<u>Carte n°3</u> Phasage par trio LDhelmet	<i>C.intestinalis</i>	0.13	0.005	-0.03*
	<i>C.robusta</i>	0.03	0.04	0.02
<u>Carte n°4</u> Phasage statistique LDhelmet	<i>C.intestinalis</i>	0.07	0.01	-0.003
	<i>C.robusta</i>	-0.04	-0.08*	-0.04*

Table des coefficients de Pearson pour les corrélation avec le taux de GC
 (* : corrélations significatives au risque alpha de 5 %)

Annexe 7B

38

Quelle carte est la plus fiable ? - Critère du taux de GC

	Echelle	1Mb	100kb	10kb
<u>Carte n°1</u> Phasage par trio LDhat	<i>C.intestinalis</i>	0.19	0.31*	0.25*
	<i>C.robusta</i>	0.02	-0.04	-0.09
<u>Carte n°2</u> Phasage statistique LDhat	<i>C.intestinalis</i>	0.37*	0.30*	0.22*
	<i>C.robusta</i>	0.10	-0.13*	-0.19*
<u>Carte n°3</u> Phasage par trio LDhelmet	<i>C.intestinalis</i>	0.13	0.005	-0.03*
	<i>C.robusta</i>	0.03	0.04	0.02
<u>Carte n°4</u> Phasage statistique LDhelmet	<i>C.intestinalis</i>	0.07	0.01	-0.003
	<i>C.robusta</i>	-0.04	-0.08*	-0.04*

Table des coefficients de Pearson pour les corrélations avec le taux de GC
 (* : corrélations significatives au risque alpha de 5 %)

Annexe 8A

39

Quelle carte est la plus fiable ? - Critère de la longueur des chromosomes

	Espèce	Coefficient de Pearson
<u>Carte n°1</u> Phasage par trio LDhat	<i>C.intestinalis</i>	0.45
	<i>C.robusta</i>	-0.51
<u>Carte n°2</u> Phasage statistique LDhat	<i>C.intestinalis</i>	-0.21
	<i>C.robusta</i>	-0.71*
<u>Carte n°3</u> Phasage par trio LDhelmet	<i>C.intestinalis</i>	-0.64*
	<i>C.robusta</i>	-0.75*
<u>Carte n°4</u> Phasage statistique LDhelmet	<i>C.intestinalis</i>	-0.72*
	<i>C.robusta</i>	-0.63*

Table des coefficients de Pearson pour les corrélations avec la longueur des chromosomes
(* : corrélations significatives au risque alpha de 5 %)

Annexe 8B

Quelle carte est la plus fiable ? - Critère de la longueur des chromosomes

	Espèce	Coefficient de Pearson	
<u>Carte n°1</u> Phasage par trio LDhat	<i>C.intestinalis</i>	0.45	
	<i>C.robusta</i>	-0.51	
<u>Carte n°2</u> Phasage statistique LDhat	<i>C.intestinalis</i>	-0.21	
	<i>C.robusta</i>	-0.71*	
<u>Carte n°3</u> Phasage par trio LDhelmet	<i>C.intestinalis</i>	-0.64*	←
	<i>C.robusta</i>	-0.75*	←
<u>Carte n°4</u> Phasage statistique LDhelmet	<i>C.intestinalis</i>	-0.72*	←
	<i>C.robusta</i>	-0.63*	←

Table des coefficients de Pearson pour les corrélations avec la longueur des chromosomes
 (* : corrélations significatives au risque alpha de 5 %)

Annexe 9

41

Le rapport r/μ chez la Cione

species	Group	GenomeSizeMb	NbChrom	GeneticSize_cM	COrate_cM_Mb	Mut_per_bp_generation	RefMu	ratio_r_u
Mus_Musculus	Animals	2448.9	19	1381.0	0.6	5.4e-09	Long2018	1.0
Homo_sapiens	Animals	2771.0	22	3594.0	1.3	1.1e-08	Long2018	1.2
Drosophila_melanogaster	Animals	133.0	4	287.3	2.2	5.5e-09	Long2018	3.9
Oryza_sativa	Plants	362.9	12	1518.0	4.2	8.2e-09	Long2018	5.1
Arabidopsis_thaliana	Plants	123.5	5	422.5	3.4	5.9e-09	Long2018	5.8
Ficedula_albicollis	Animals	979.5	29	3014.0	3.1	4.6e-09	Long2018	6.7
Caenorhabditis_elegans	Animals	78.3	5	250.0	3.2	2.7e-09	Long2018	11.8
Daphnia_pulex	Animals	197.2	12	1206.0	6.1	4.0e-09	Long2018	15.3
Apis_mellifera	Animals	229.8	16	4115.0	17.9	5.0e-09	(1)	35.8
Ciona_intestinalis	Animals	123.0	14	4702.9	38.2	5.0e-09	(1)	76.4
Chlamydomonas_reinhardtii	Plants	120.4	17	1000.0	8.3	9.4e-10	Long2018	88.5
Plasmodium_falciparum	SAR	23.3	14	1630.4	70.1	2.4e-10	Long2018	2861.5
Schizosaccharomyces_Pombe	Fungi	12.6	3	2220.0	176.8	2.0e-10	Long2018	8837.6
Saccharomyces_cerevisiae	Fungi	12.1	16	4420.0	366.2	1.7e-10	Long2018	21928.0
Paramecium_tetraurelia	SAR	76.8	162	30543.6	397.7	2.0e-11	Sung2012	198832.4

Distribution du rapport r/μ chez différentes espèces eucaryotes