Whole Genome Sequence of the European Bison *a window to the genomic history of bovine domestication*

Laurence Flori, Madeyska Anna, Hugues Parinello, Hubert Leveziel and Mathieu Gautier

16 décembre 2014

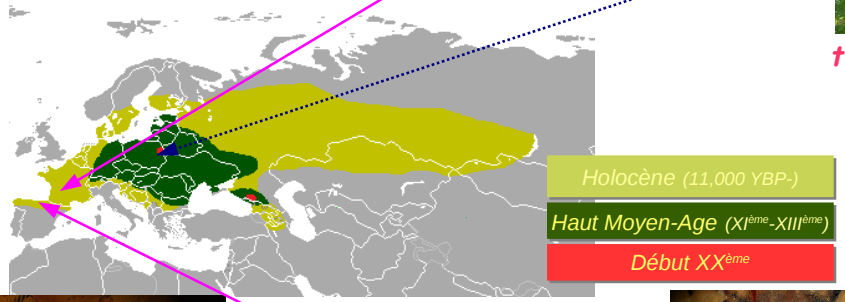
Le bison européen *(Bison bonasus)* comme modèle d'étude

- Animal emblématique de la faune sauvage européenne :
  - Plus gros mammifère terrestre vivant
  - Plus proche parent sauvage des bovins domestiques *(Bos taurus)*

Histoire du bison européen

†*Auroch (B.primigenus,?-1627)*

Grotte de Lascaux (18000 YBP)



Grotte de Altamira (15000 YBP)

Principaux objectifs de l'étude

Questions

- Caractérisation fine de l'histoire démographique du Bison Européen
  - Variation de l'effectif efficace (N_e) au cours du temps
  - Lien avec l'espèce bovine
- Comparaison sur une échelle pan-génomique des gènes soumis à sélection entre le bison et le bovin domestique
  - Identification de cibles potentielles de la domestication

Plan d'expérience

Matériel d'étude

- Séquençage complet du génome (WGS) de deux bisons européens à couverture moyenne (10-15X)
- Ressources génomiques disponibles sur les espèces modèles bovine (génome taurin assemblé et annoté + zébu + 1,000 Bull Genome project) et ovine (dans une moindre mesure)

Méthodes d'analyses

- Caractérisation de la diversité génétique
- Application de diverses méthodes d'inférence démographique à partir de génome complet
- Recherche de gènes sous sélection par comparaisons de séquence (KaKs)

Séquençage

- Echantillonnage (A. Madeyska-Lewandowska, H. Leveziel)
  - 2 mâles : BBO_3569 et son fils BBO_3574 (validé a posteriori)
  - Prélevés en 1991 dans la forêt primaire de Biolowieza (Pologne)
- Séquençage sur deux pistes de HiSeq2000
  - plateforme MGX (Montpellier)
- Mapping des reads sur le génome bovin (pas d'assemblage *de novo*)
  - Reference :
    - assemblage UMD3.1 (BTA1–BTA29, BTAX, BTAmt)
    - assemblage Btau7 (BTAY)
  - *bwa* : options par défaut retenues (pas d'influence notable)
  - *samtools* :
    - rmdup ; -q 20
    - mpileup : "prob. realignment" (GATK indel-realigner ou –B : pas d'influence)

Couverture (bilan)

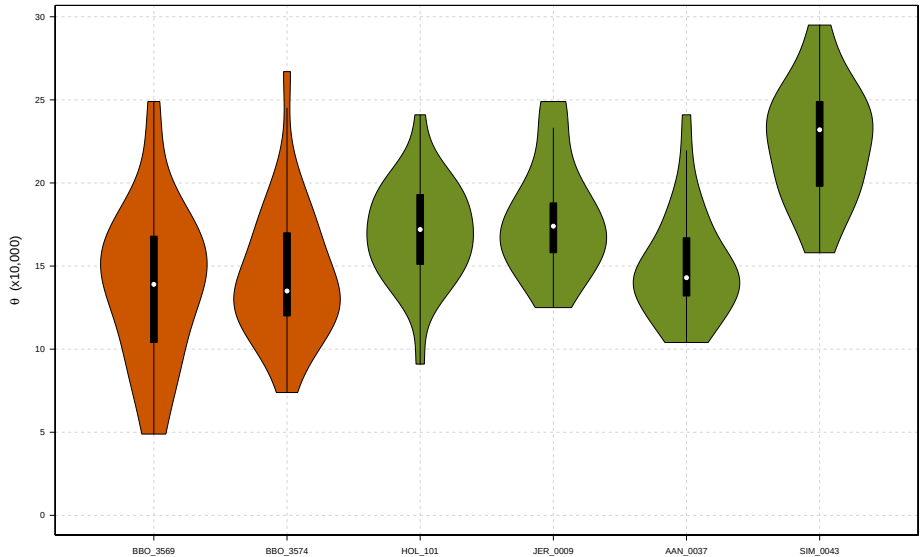
BTA	Size (in bp)	Coverage (% of sequence covered)	
		BBO_3569	BBO_3574
All autosomes	2,512,082,506	9.81 (95.5)	11.6 (95.7)
BTAX (including PAR)	148,823,899	5.57 (90.9)	6.48 (91.2)
BTAY$^{(*)}$	43,300,181	66.3 (12.2)	76.8 (12.4)
BTAmt	16,338	397 (97.7)	302 (97.1)

($*$) Gros déficit de lectures correctement pairées + forte similarité avec BTAX : pb assemblage BTAY ?

Hétérozygotie (*mlRho* Haubold et al. (2010) ; Lynch (2009))

Individu	couverture	nsites	$\hat{\theta} = 4N_e\mu$ (x10,000)	$\hat{\rho} = 4N_e c$	$\hat{\epsilon}$ (%)
BBO_3569	3<DP<100	2,263,075,937	14.0	$1.28\text{x}10^{-6}$	0.155
	10<DP<100	1,159,212,119	13.9	$1.26\text{x}10^{-6}$	0.144
BBO_3574	3<DP<100	2,301,336,331	14.7	$1.36\text{x}10^{-6}$	0.227
	10<DP<100	1,512,223,055	14.0	$1.10\text{x}10^{-6}$	0.214
AAN_0037	10<DP<100	572,695,652	14.4	$5.86\text{x}10^{-6}$	0.240
JER_0009	10<DP<100	1,129,319,115	17.5	$4.12\text{x}10^{-6}$	0.106
HOL_0101	10<DP<100	1,006,216,760	16.8	$5.29\text{x}10^{-6}$	0.251
SIM_0043	10<DP<100	814,576,967	22.1	$8.51\text{x}10^{-6}$	0.139

θ par chromosome



Comparaison avec HSA (Prufer et al., 2014)

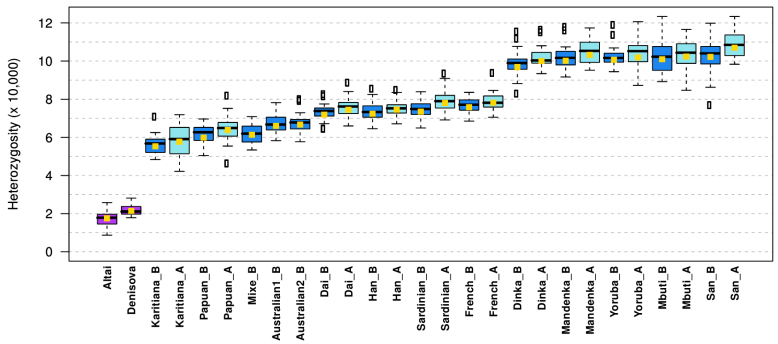


Figure S9.1: Heterozygosity estimates and autosomal distributions. The yellow squares are the average genomic heterozygosity estimates (i.e. those in Table S9.1) for 10,000 sites and the box-and-whiskers represent the distributions across 22 autosomes given by the R-function 'boxplot' with default parameters[3]. Archaic samples are in purple; modern human from A- and B-panels are in light and dark blue, respectively. The samples 'Australian1_B' and 'Australian2_B' are 'WON,M' and 'BUR,E', respectively (see Table S9.1).

Le modèle PSMC (Li et Durbin, 2011)

LETTER

doi:10.1038/nature10231

Inference of human population history from individual whole-genome sequences

Heng Li[1,2] & Richard Durbin[1]

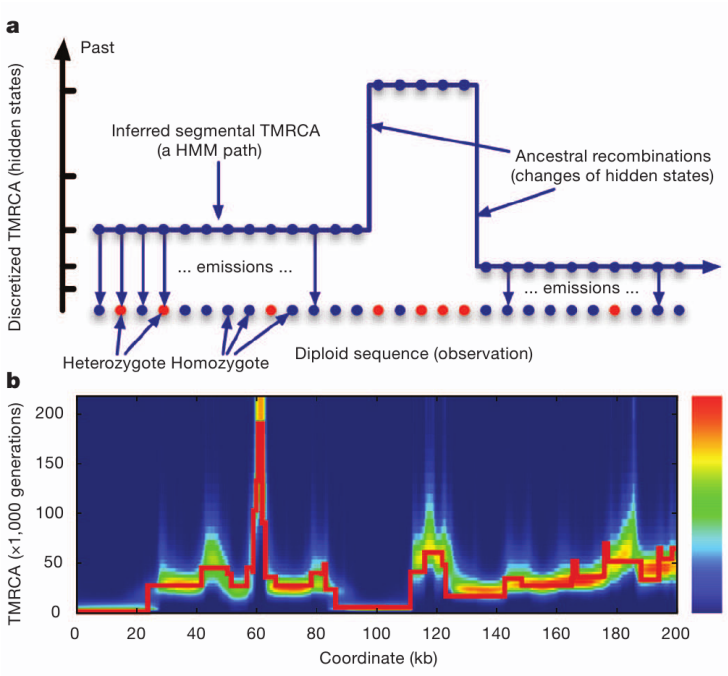


Figure 1 | Illustration of the PSMC model and its application to simulated data. a, The PSMC infers the local time to the most recent common ancestor (TMRCA) on the basis of the local density of heterozygotes, using a hidden Markov model in which the observation is a diploid sequence, the hidden states are discretized TMRCA and the transitions represent ancestral recombination events. **b**, We used the ms software to simulate the TMRCA relating the two alleles of an individual across a 200-kb region (the thick red line), and inferred the local TMRCA at each locus using the PSMC (the heat map). The inference usually includes the correct time, with the greatest errors at transition points.

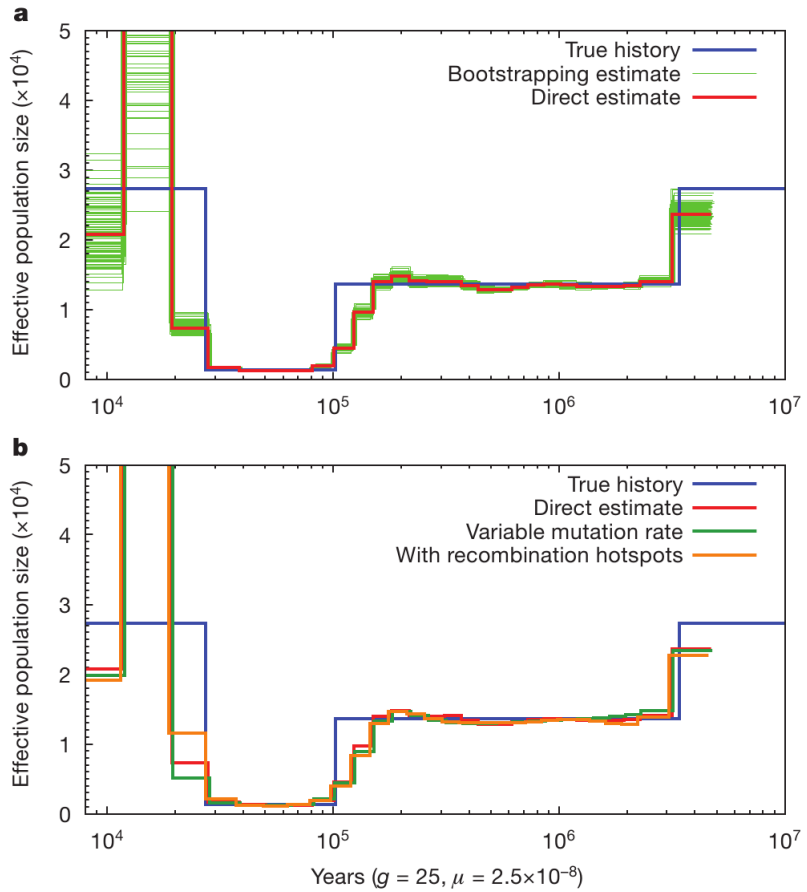
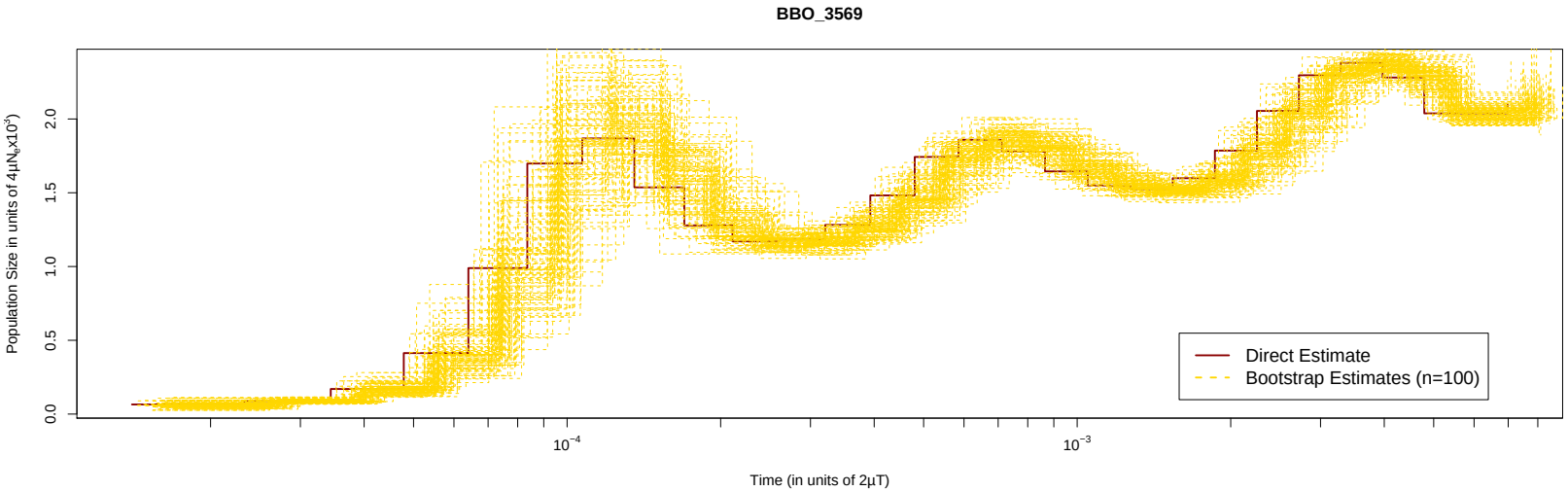
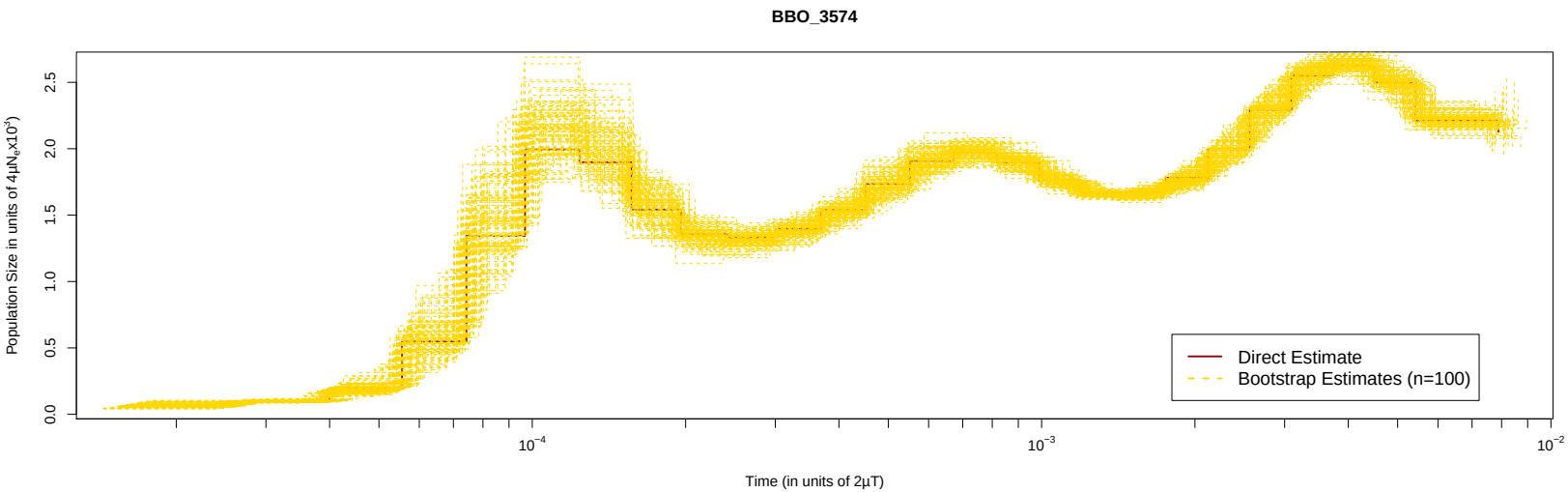


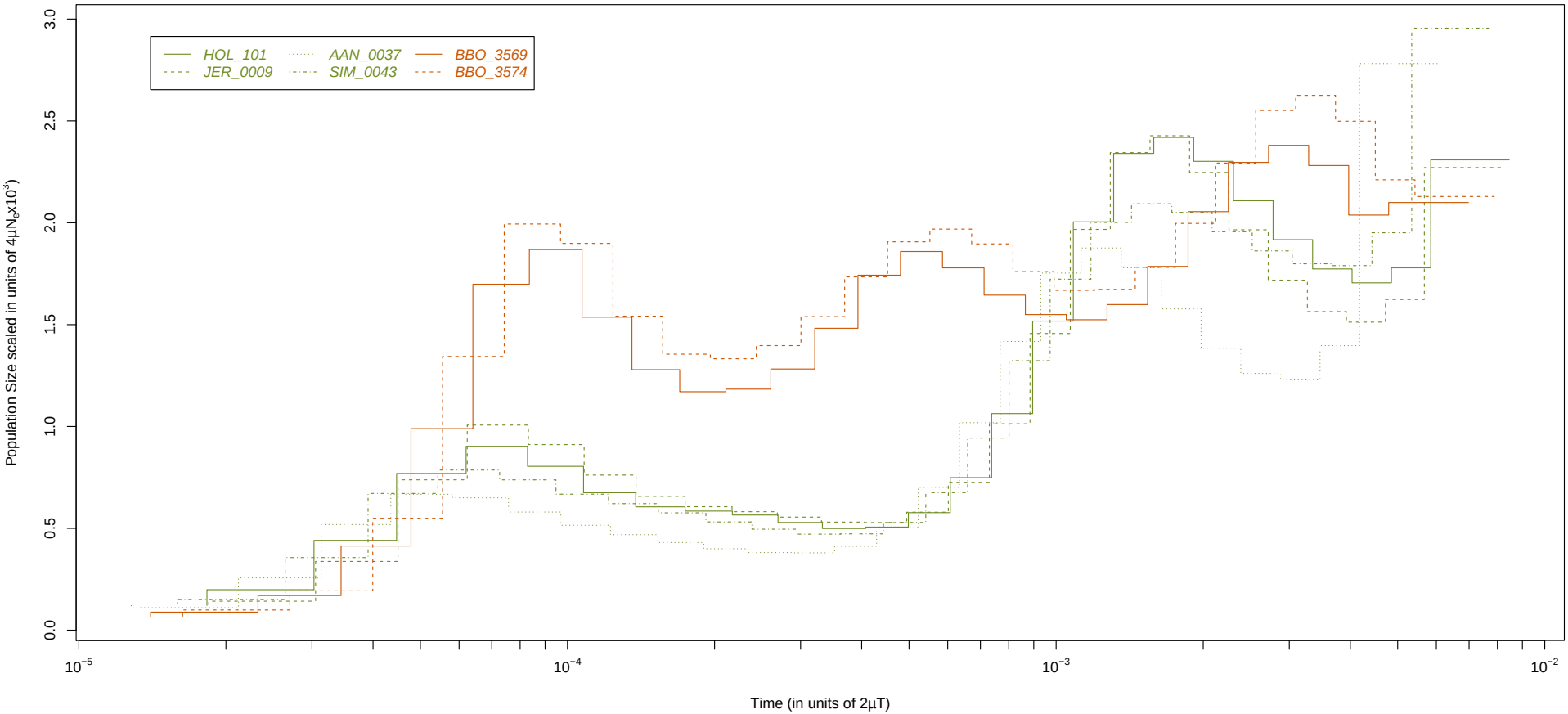
Figure 2 | PSMC estimate on simulated data. a, PSMC estimate on data

Histoire démographique du Bison : analyse PSMC (Li et Durbin, 2011)





Histoire démographique du Bison : analyse PSMC (Li et Durbin, 2011)



Histoire démographique du Bison : biais ?

Calling *SAMtools* sur génome divergent

- Biais vers allèle de l'assemblage de référence (Nevado et al., 2014)
- Illustration avec erreurs de génotypages (Père et Fils homozygotes à des allèles différents)

couverture	DP>5	DP>10	DP>15	DP>20	DP>25	DP>30
nsites	2,650,252	876,350	104,046	29,424	13,816	6,782
ϵ_{geno} en %	11.4	12.7	5.90	0.360	0.0072	0.00

Conséquences ?

- Augmentation du taux de mutation apparent
- Surestimation de t et N_e

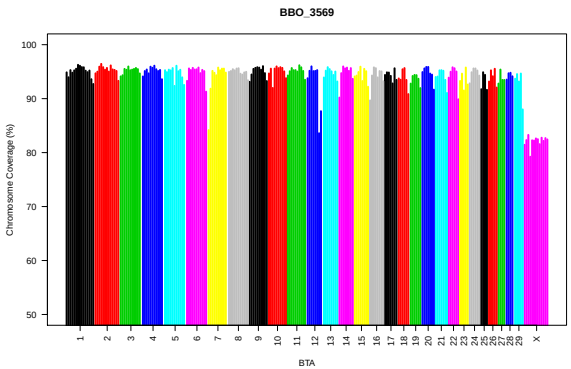
Solution ? (en cours de developpement)

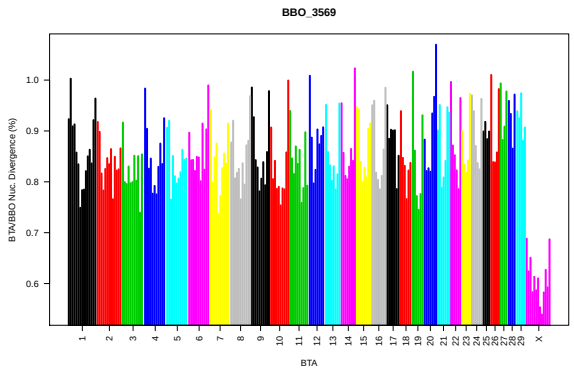
- Calling sans génome de référence

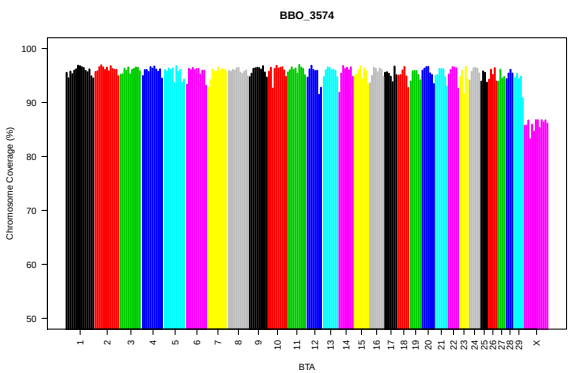
Divergence BBO/BTA

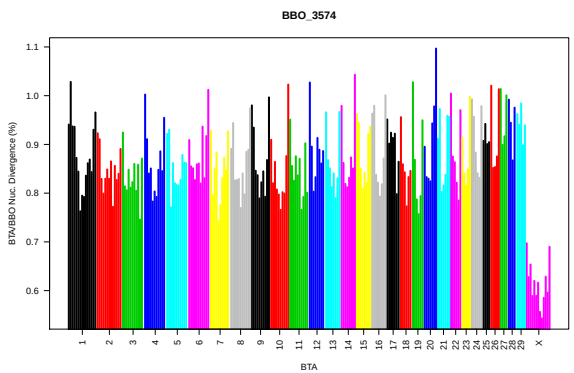
Définition de séquences BBO consensus

- BAQ>25 ; 3<DP<100 ; no indels (Déf. de "chunks" de ≤10 Mb ; 256 A, 14 X, 1MT)
- Ref. Allèle = allèle majoritaire (lectures)



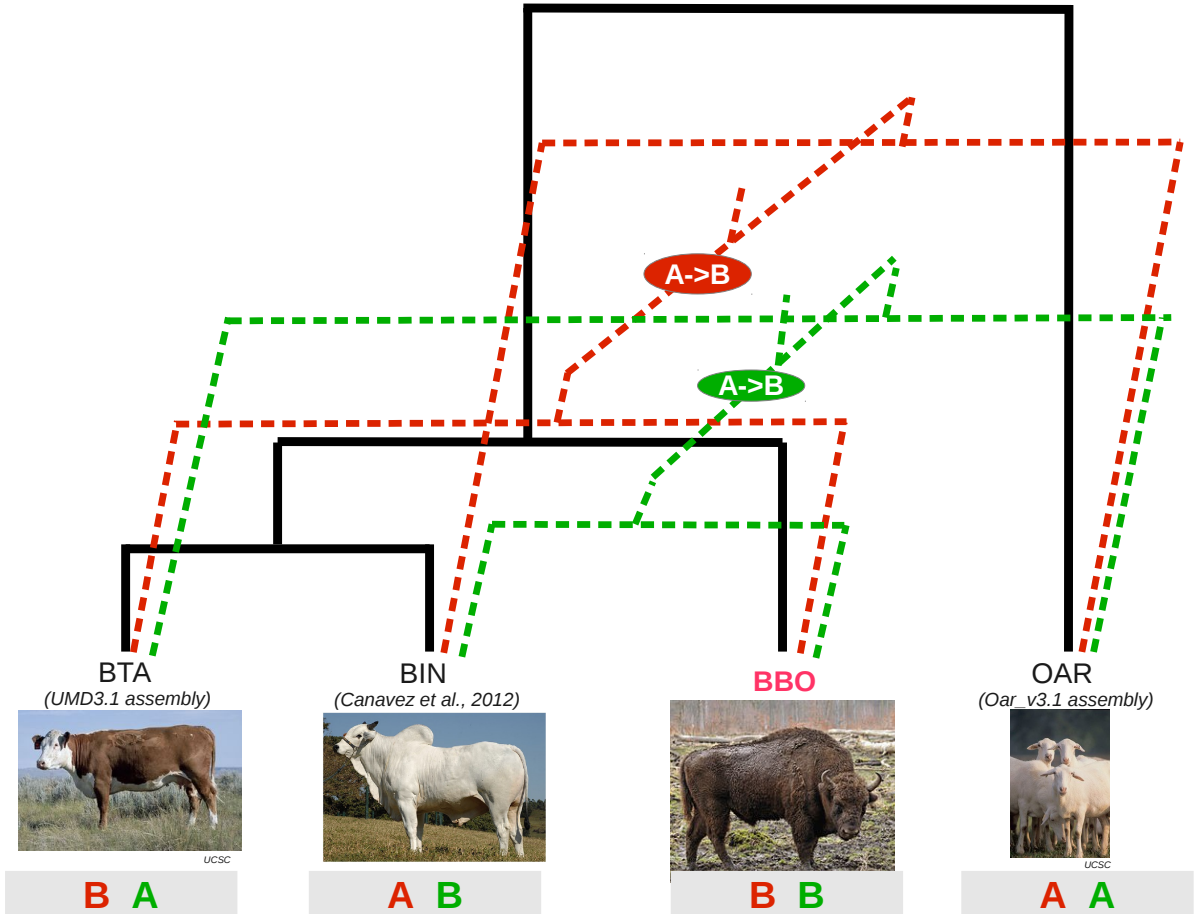






Divergence (%)	BBO_3569	BBO_3574
Auto.	0.869	0.880
BTAX	0.634	0.639
BTAmt	3.58	3.41

Admixture Bovin/Bison européen ? (Green et al., 2010 ; Durand et al., 2011)



nsnp BABA	nsnp ABBA	Dstat (x10,000)
3,611	3,604	9.70×10^{-4} (n.s)

Datation de la divergence BTA/BBO : Méthode 1 reich et al. (2013)

Principe : $\mathrm{P}(A \mid B)$ décroît avec t_D (e.g. $\mathrm{P}_{t=0}(A \mid B) \simeq \frac{1}{3}$)

- A=random chromosome (ici ref. assembly) de l'espèce A porte un allèle dérivé
- B= site hétéro. dans un individu de l'espèce B (ici BBO)

Observations (pour chaque individu)

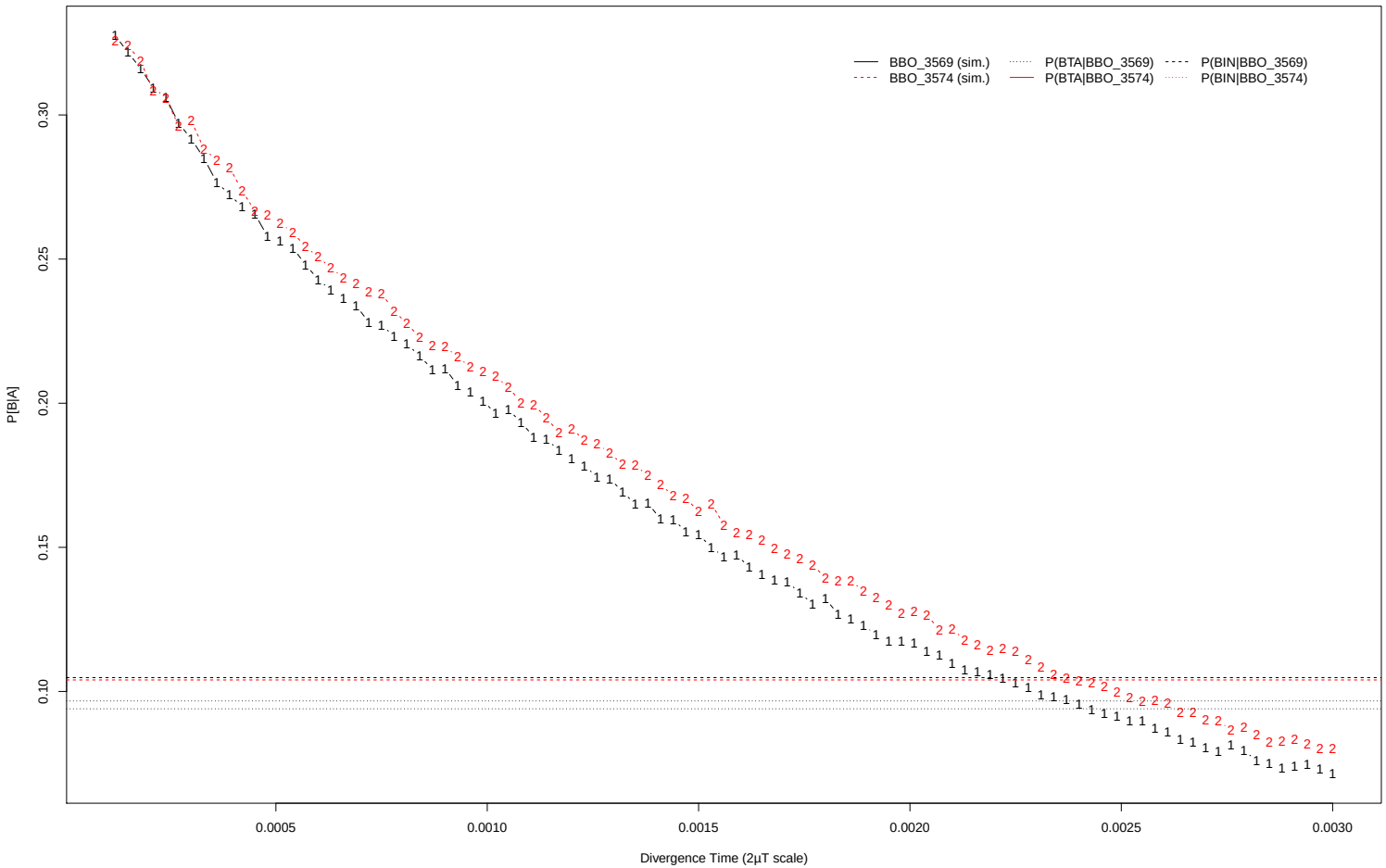
- Identification des sites hétéros dans BBI (bcftools ; 10<DP<100 ;Q>20)
- Récupération de l'allèle de référence (BTA et BIN) et polarisation (OAR)

$\widehat{\mathrm{P}}(A \mid B)$	A=BTA	A=BIN
B=BBO_3569	10.48% (n=312,656)	9.668 % (n=262,738)
B=BBO_3574	10.43% (n=431,610)	9.640 % (n=360,547)

Calibration de t_D par simulations (selon démographie inférée par *psmc* pour B)

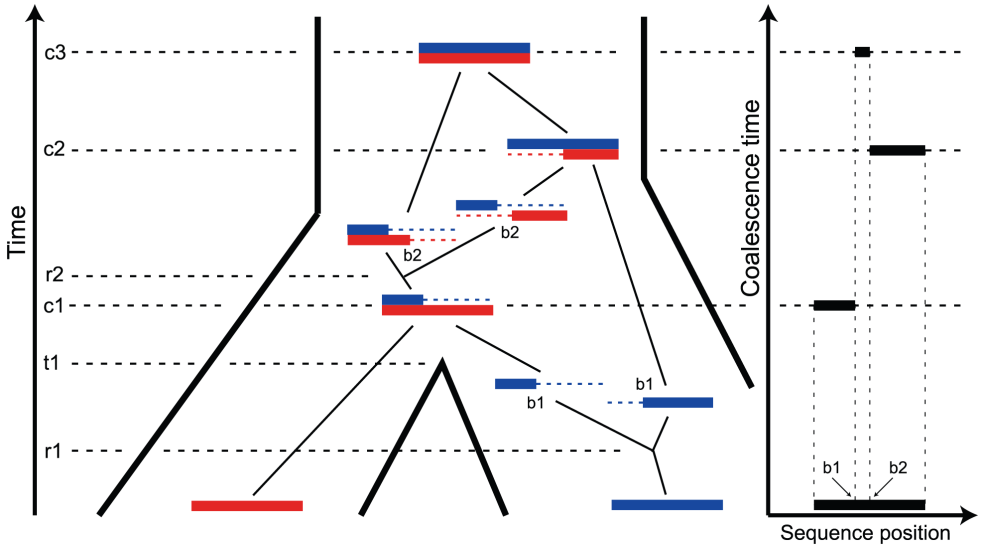
- génome simulé= 100 segments de 10 Mb

Datation de la divergence BTA/BBO : Méthode 1 reich et al. (2013)



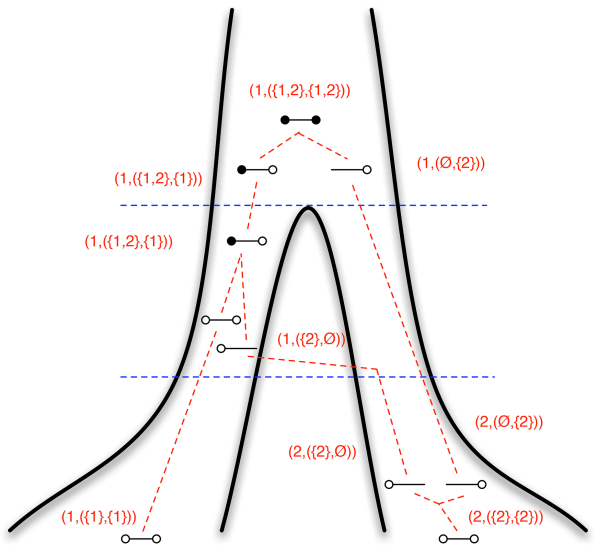
t_D	BTA | BBO_3569	BIN | BBO_3569	BTA | BBO_3574	BIN | BBO_3574
en 2μT	0.00222	0.00237	0.00237	0.00258
en YBP (μ=2.2×10^{-9})	505,000	539,000	539,000	586,000

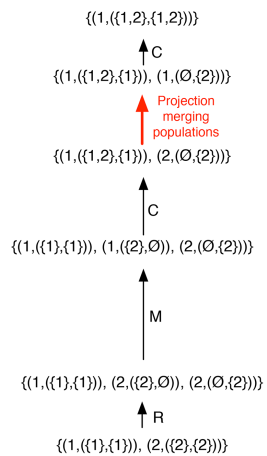
Divergence BTA/BBO : Méthode 2 (*coalhmm* Dutheil (2009), Mailund (2011,2012))



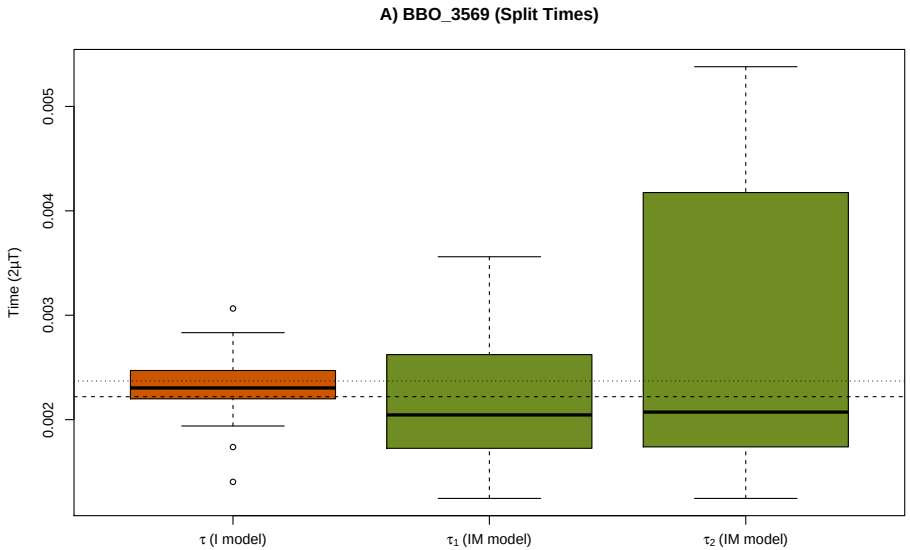
Modèle I (Mailund et al., 2011)

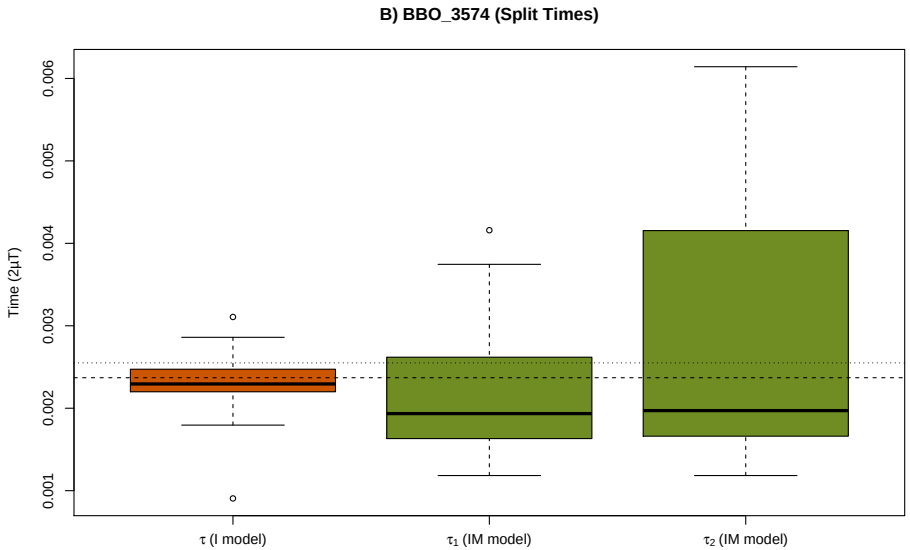
Modèle IM (Mailund et al., 2012)

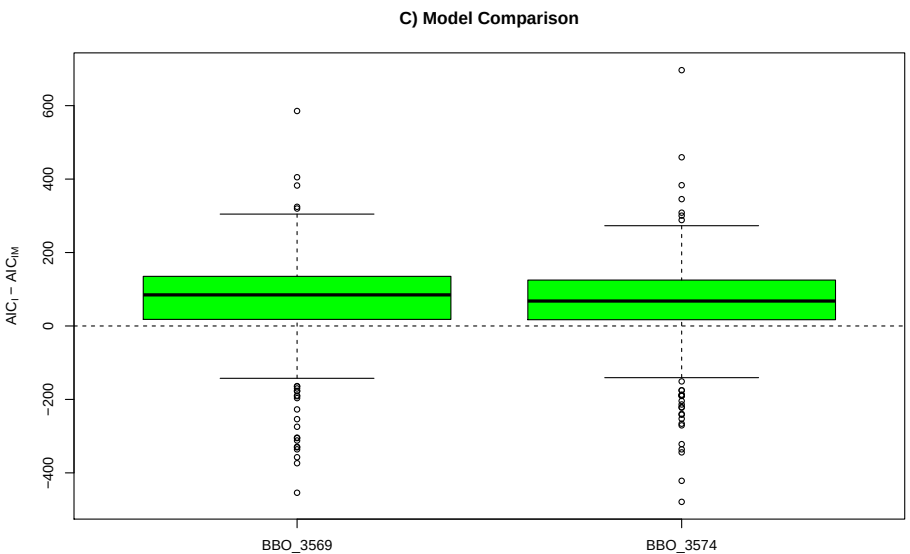


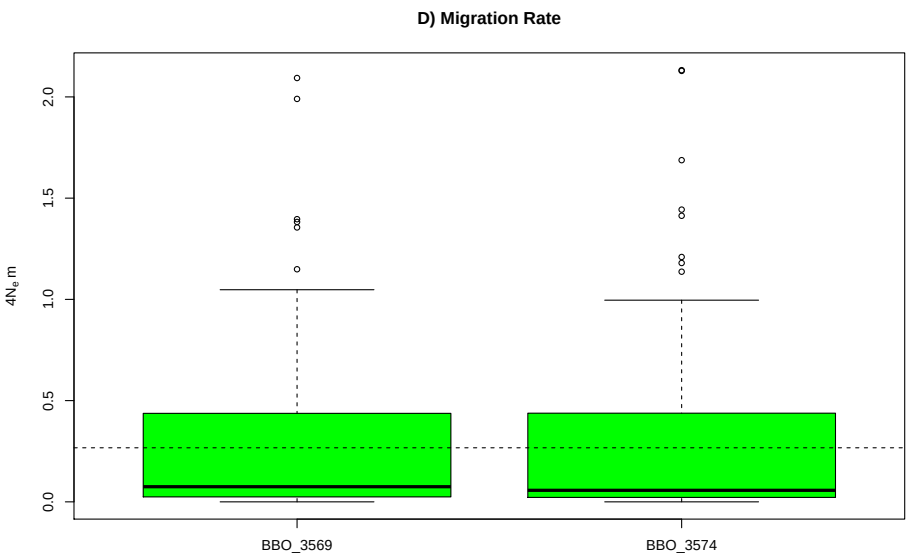


Divergence BTA/BBO : Une spéciation sympatrique

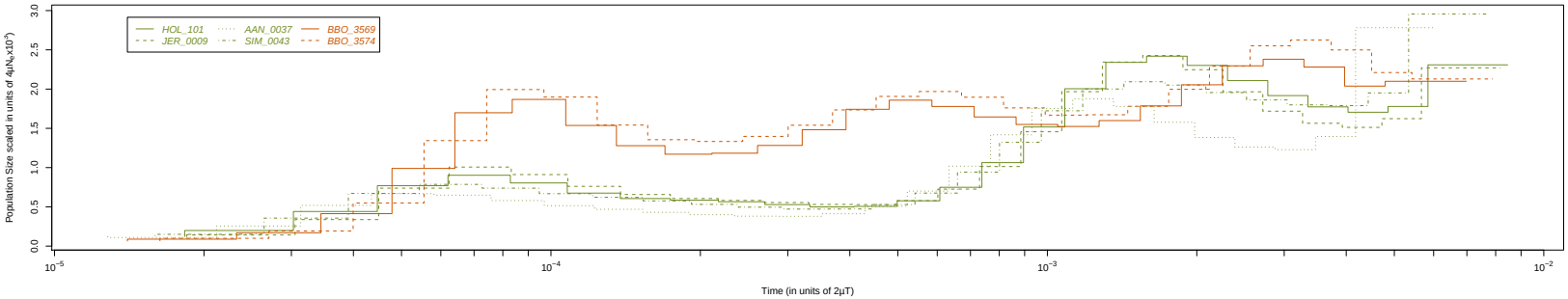


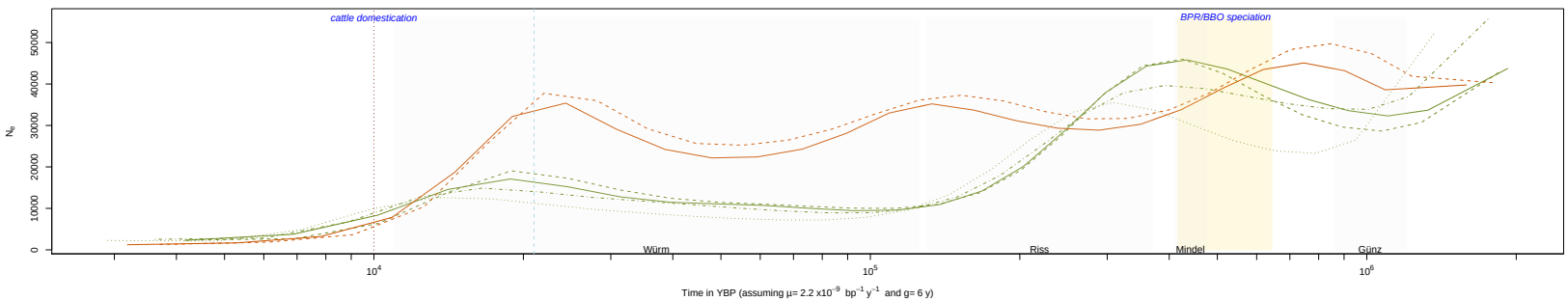


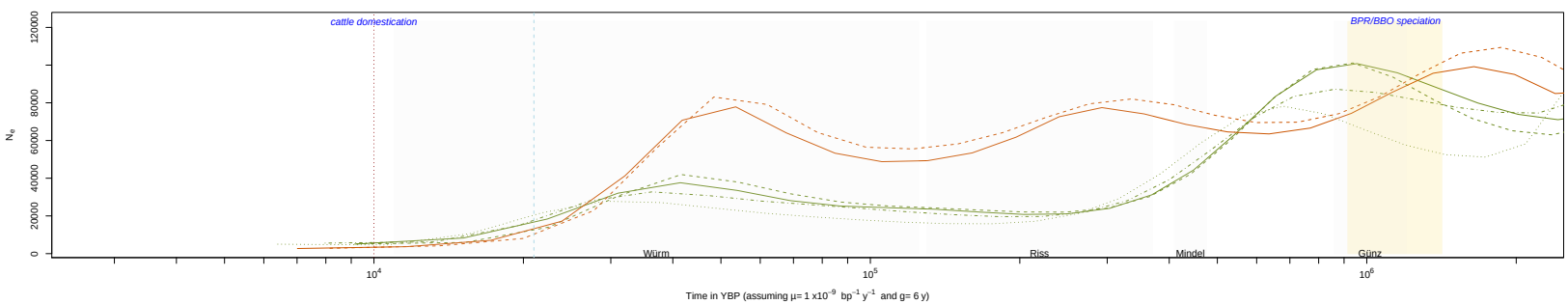




Histoire démographique du Bison





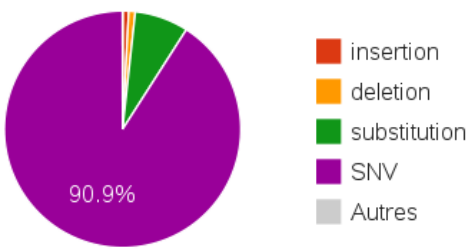


Annotation des variants (10 <DP< 100 ; Q>20 ; w>3) avec VEP (Cingolani et al., 2012)

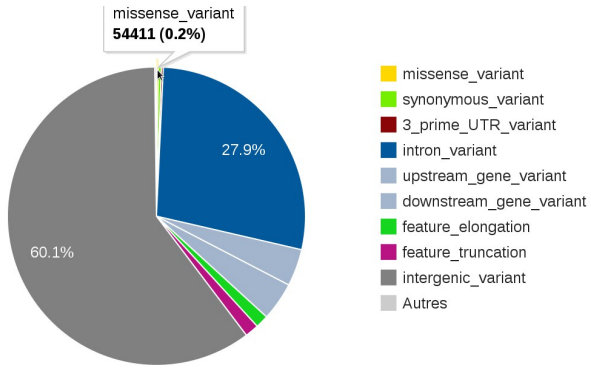
A) General Statistics

Lines of input read	24365916
Variants processed	24365886
Variants remaining after filtering	24365886
Lines of output written	26266563
Novel / existing variants	19100031 (78.4%)/ 5265855 (21.6%)
Overlapped genes	24558
Overlapped transcripts	26680
Overlapped regulatory features	-

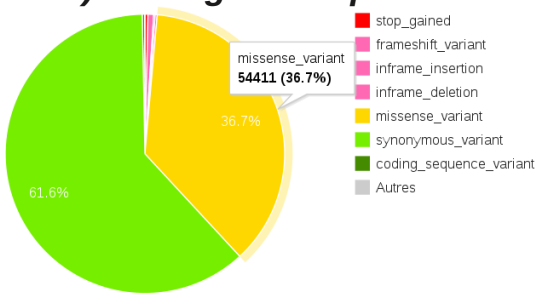
B) Variant classes



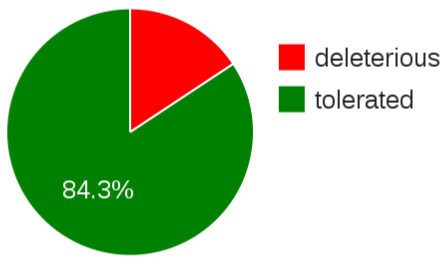
C) Consequences



D) Coding consequences



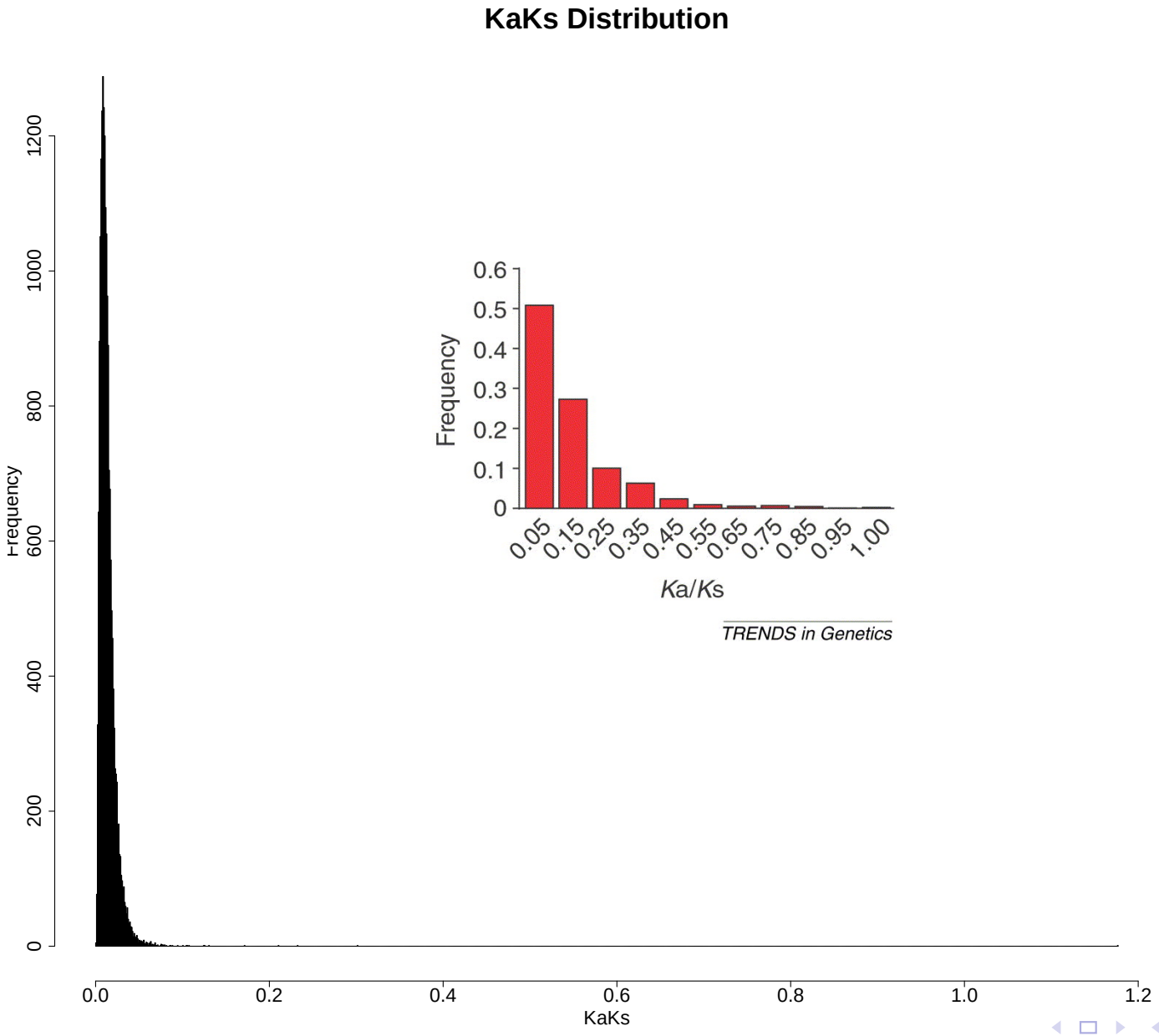
E) SIFT summary

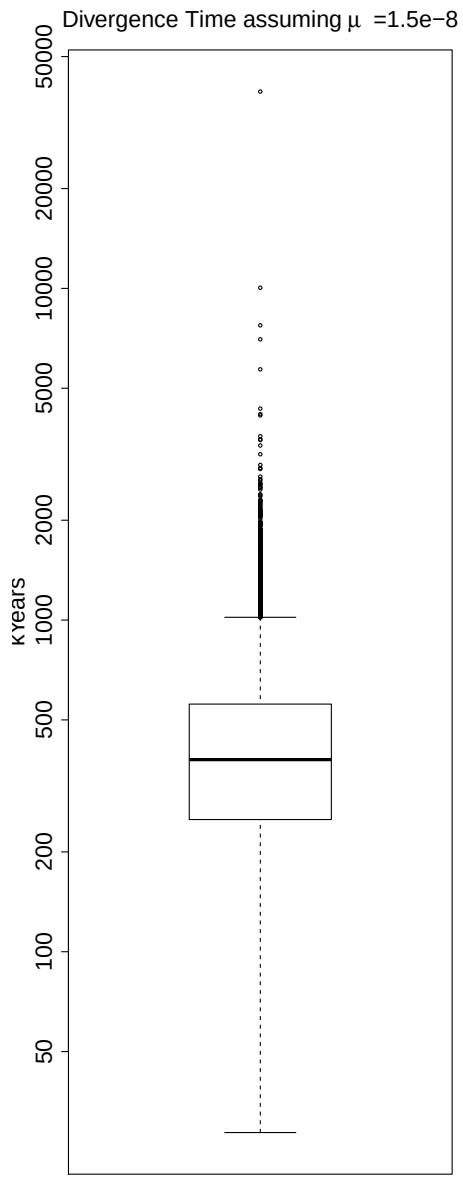


Recherche de gènes sous sélection (KaKs)

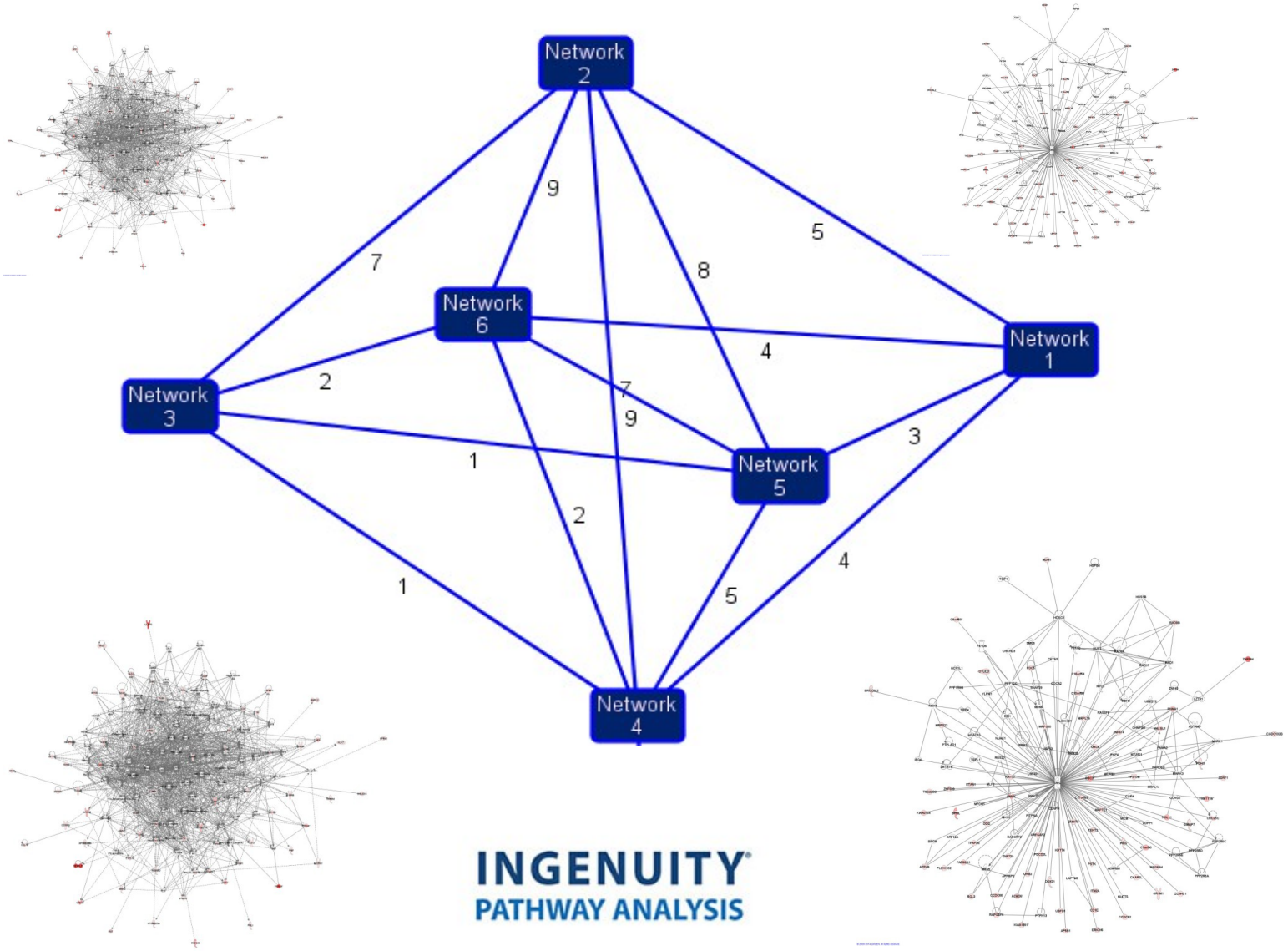
- ENSEMBL gene ID bovins et séquences consensus BBO correspondantes
- Calcul des KaKs (*KaKs Calculator* (Zhang et al., 2006)) : MLWL model (Tseng et al., 2004)
- Au total, 19,141 transcrits dont 970 avec un KaKs>1

Recherche de gènes sous sélection (KaKs)

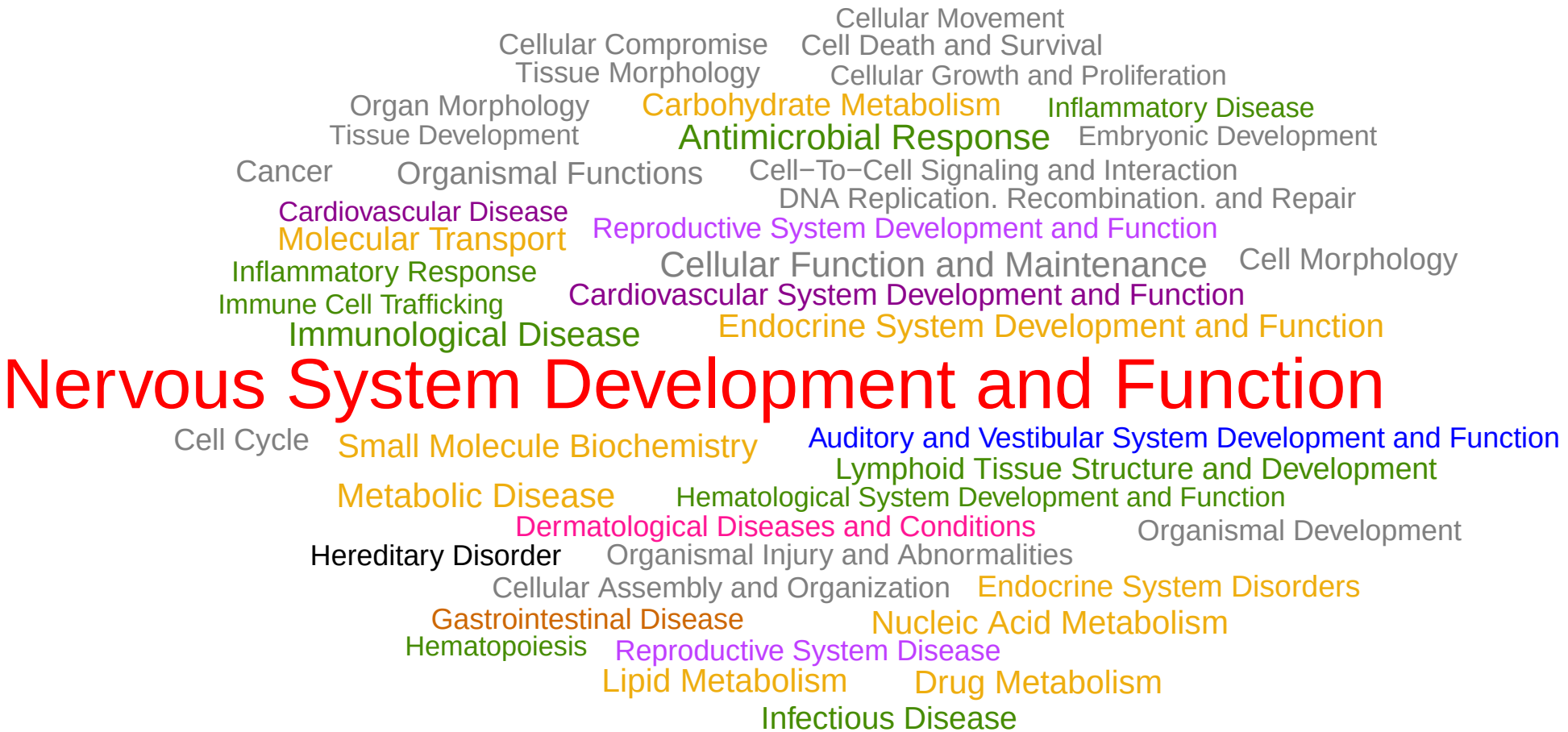




Analyse globale (IPA) : 888 gènes (KaKs>1) ; 473 annotés



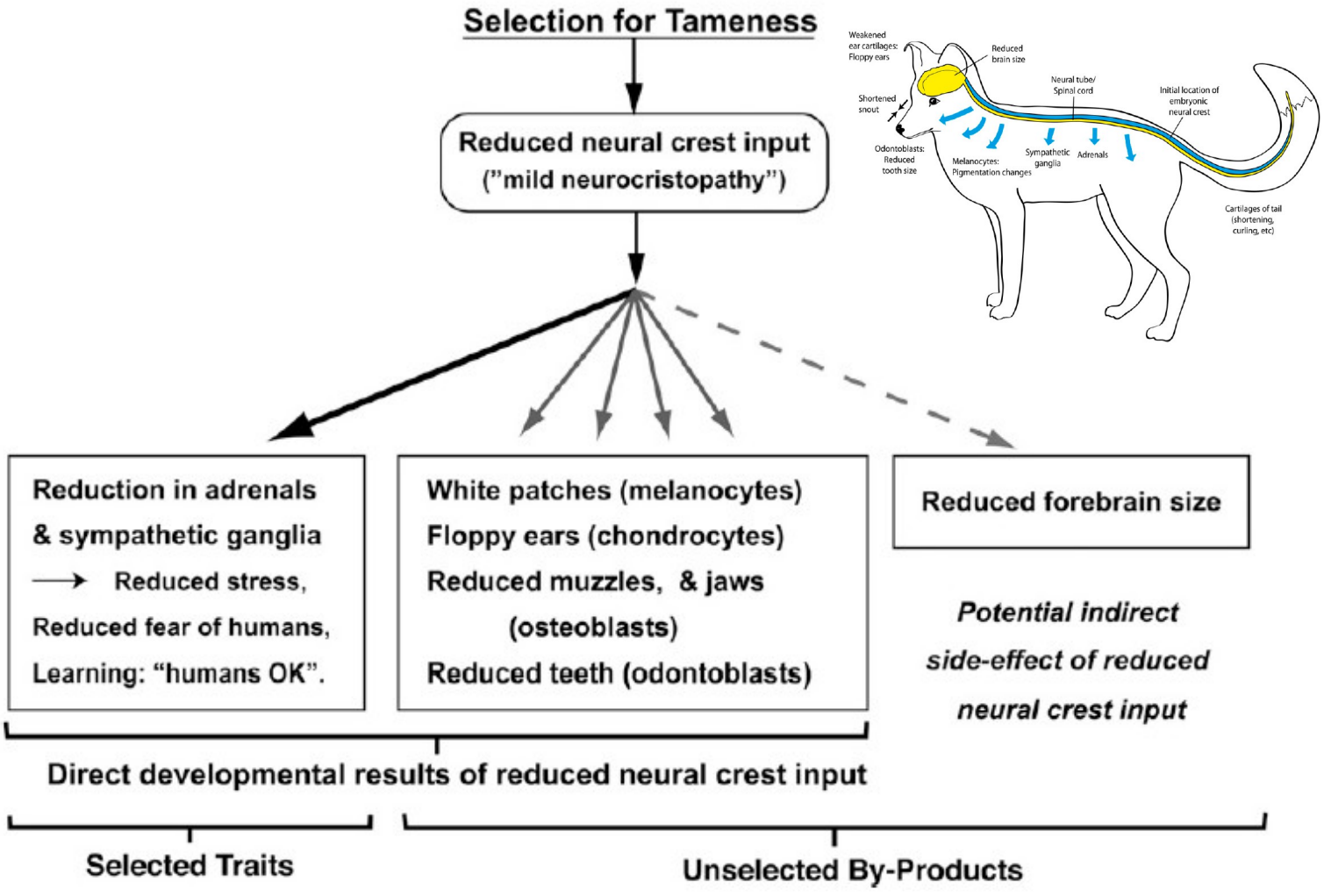
Annotation fonctionnelle (IPA)



Annotation fonctionnelle (IPA)

- Fonction la plus significative : Nervous System
  - Olfaction (e.g. Olfr), Goût (e.g. Tastr), Vision (e.g. Rét. Pigm.), Audition
  - Neurogénèse (e.g. cellules progénitrices neuronales, dendrites, astrocytes...)
- Métabolisme
  - Régulation de la glycémie, métabolisme des lipides
  - Reproduction e.g. Production et régulation de la testostérone : Hyperandrogénie
- Propriétés de la peau et des poils
  - coloration (albinisme, mélanome)
  - structure et développement des poils (whooly hair, hypotrichose)
- Développement des Dents, Os et Cartilage
- Réponse Immunitaire
  - Réponses antimicrobienne et antivirale (e.g. catégorie "Infectious Disease")
  - "Immune Cell Trafficking" ; "Inflammatory Response"
- Développement et Fonctionnement de la Glande Mammaire

Rôle central de la crête neurale dans le syndrôme de domestication (Wilkins, 2014))



Conclusions

Histoire du Bison Européen

- Séparé il y a environ 500,000 ans de l'aurochs (Spéciation sympatrique)
- Démographie marquée par les périodes glaciaires (mieux supportées que l'auroch)
- Chute démographique récente liée à l'anthropisation de son milieu

Voies physiologiques de la domestication bovine

- Accord avec l'hypothèse de Wilkins sur le rôle central de la crête neurale
- Les phénotypes "co-produits" de la domestication seraient liés aux effets pléïotropes des gènes impliqués dans le développement de la crête neurale

Techniquement

- Séquençage complet d'individus d'espèce proche d'animaux modèles (assemblage+annotation) : une grande quantité d'information (attention à certains écueils) à moindre coût (wet lab)