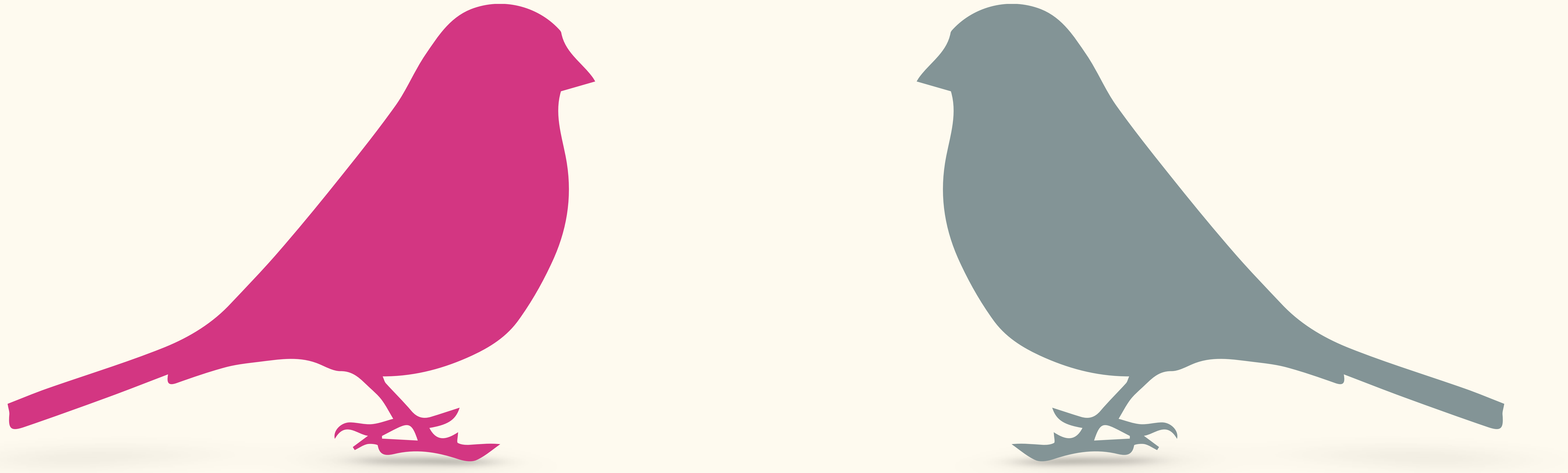
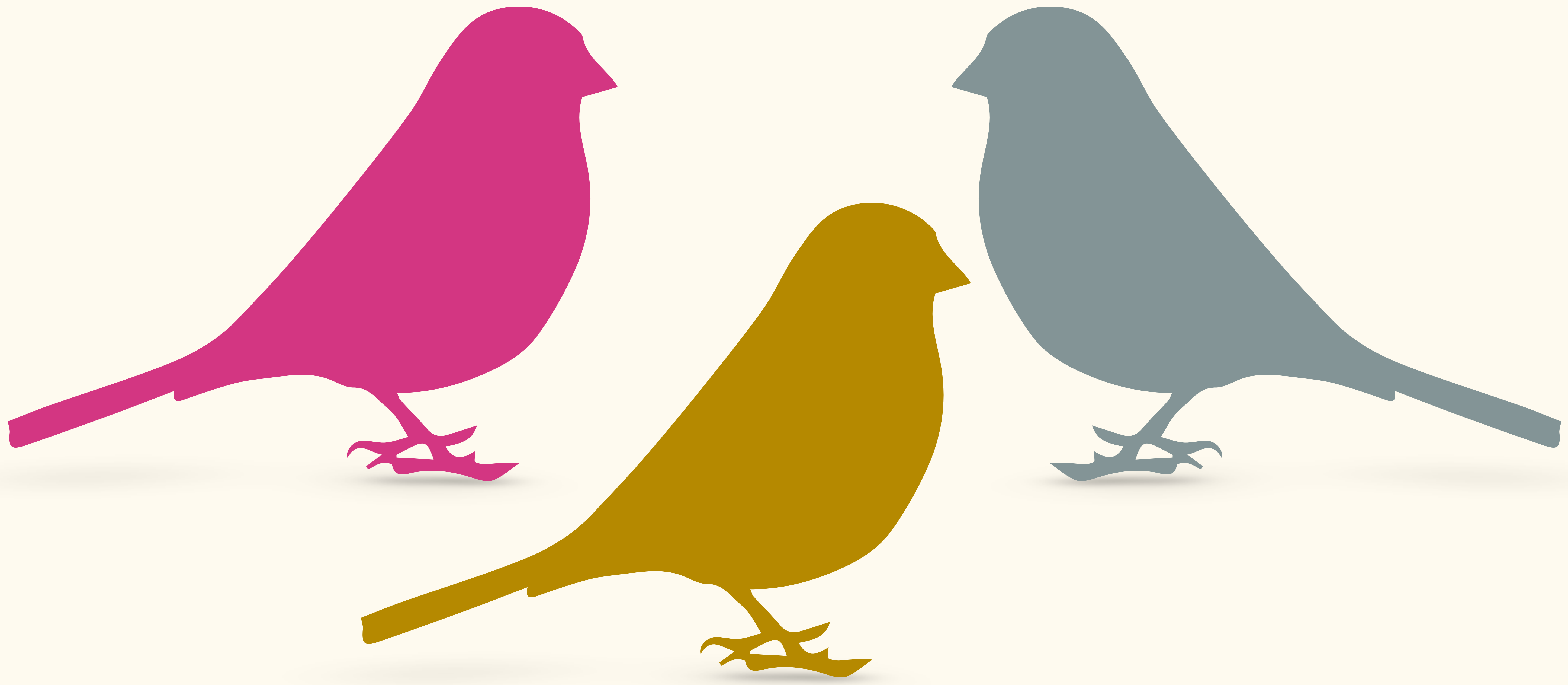
Tree-based introgression detection

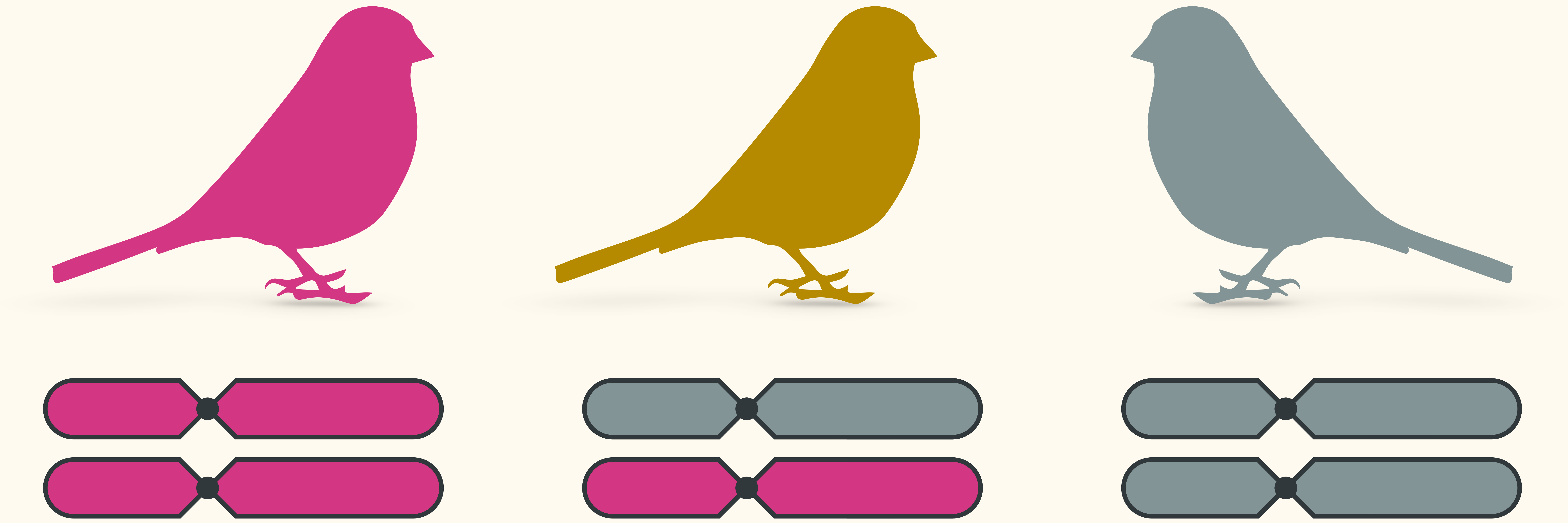
Hybridization



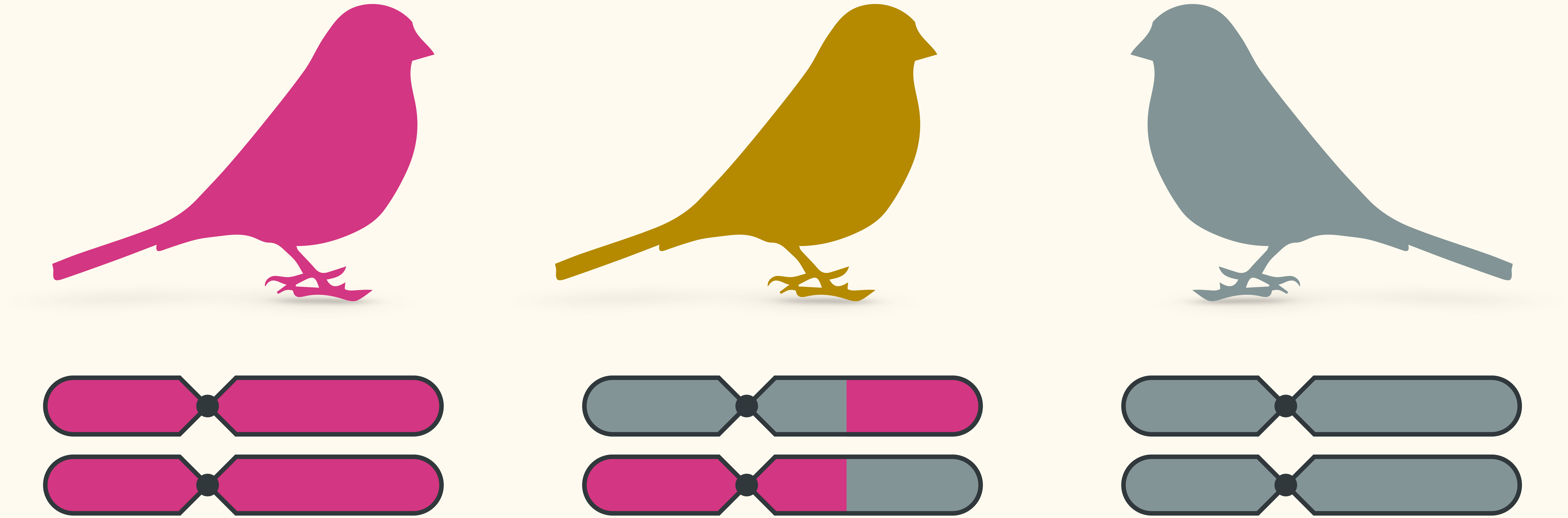
Hybridization



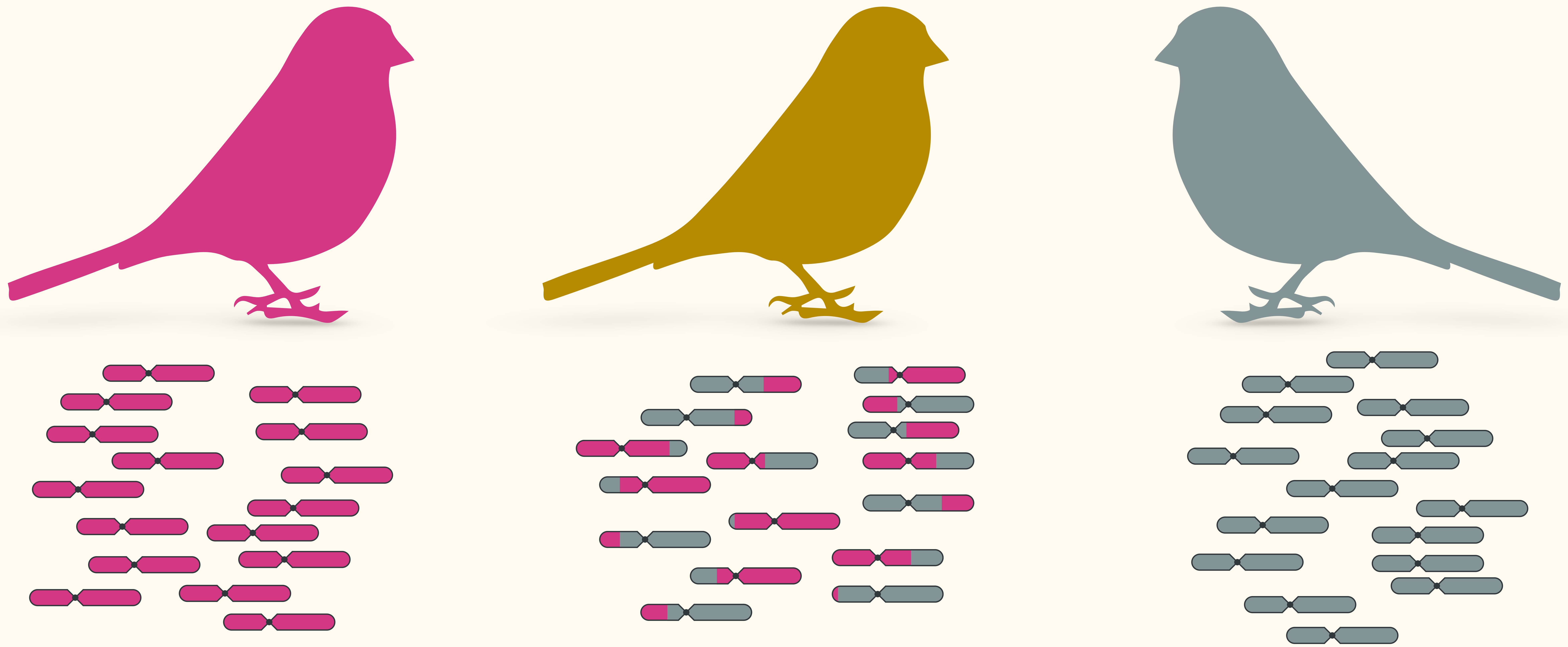
Hybridization



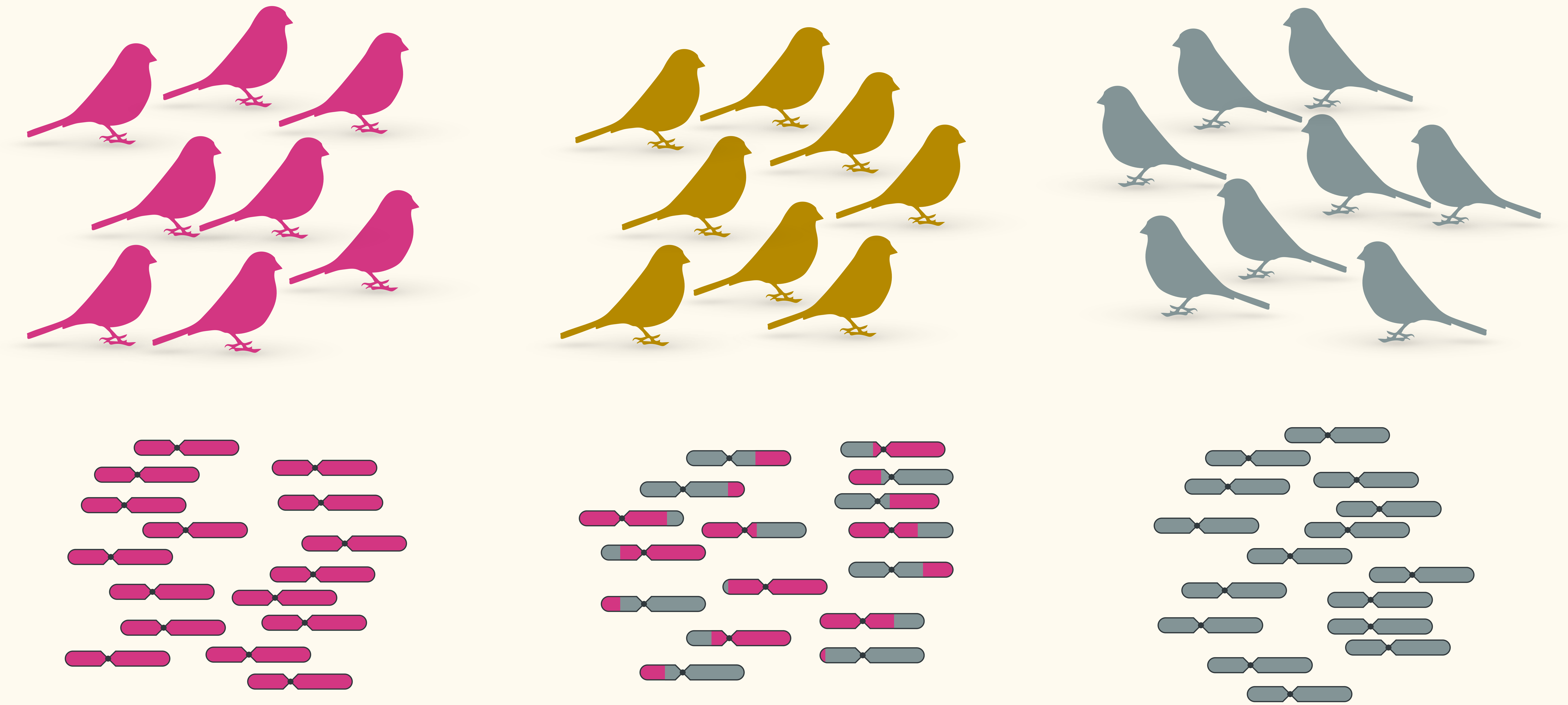
Hybridization



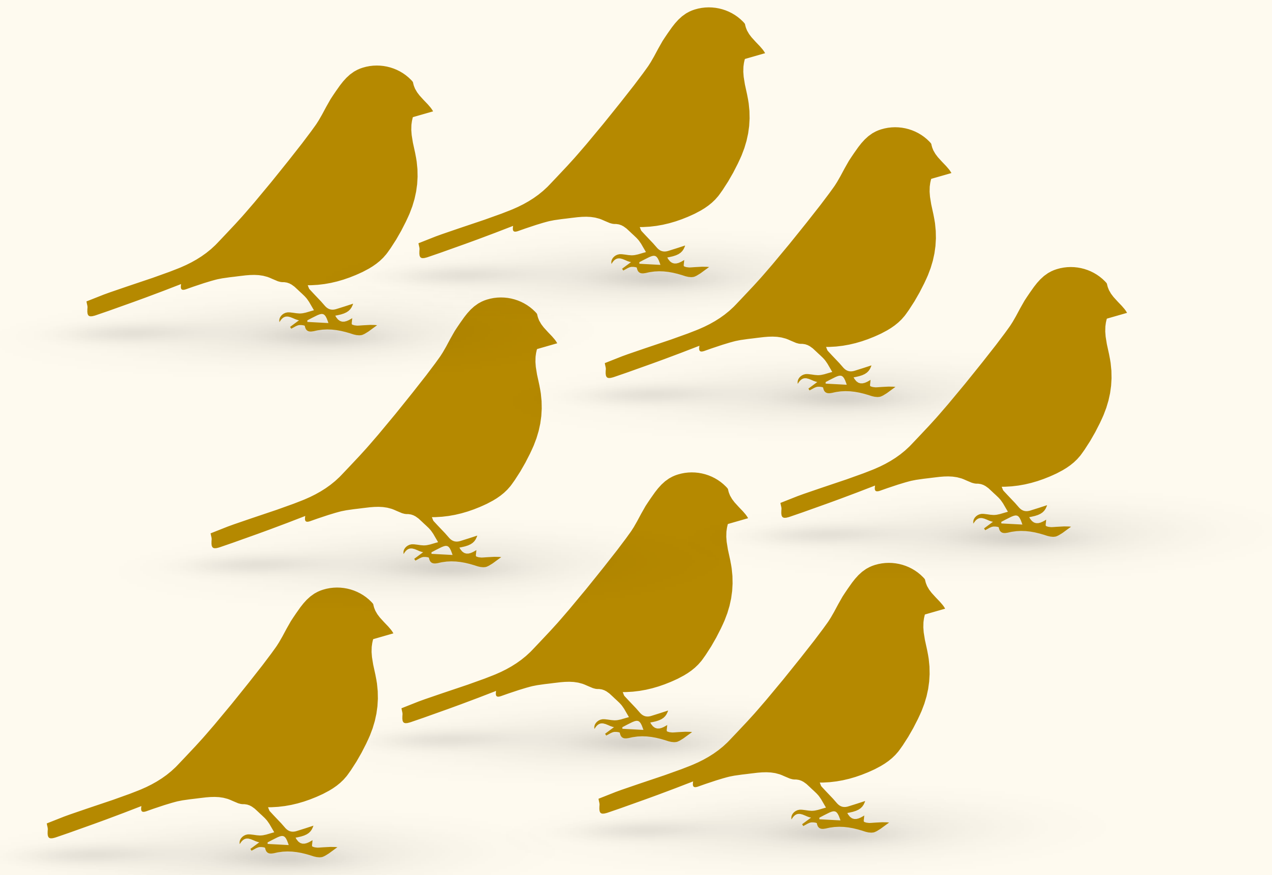
Hybridization

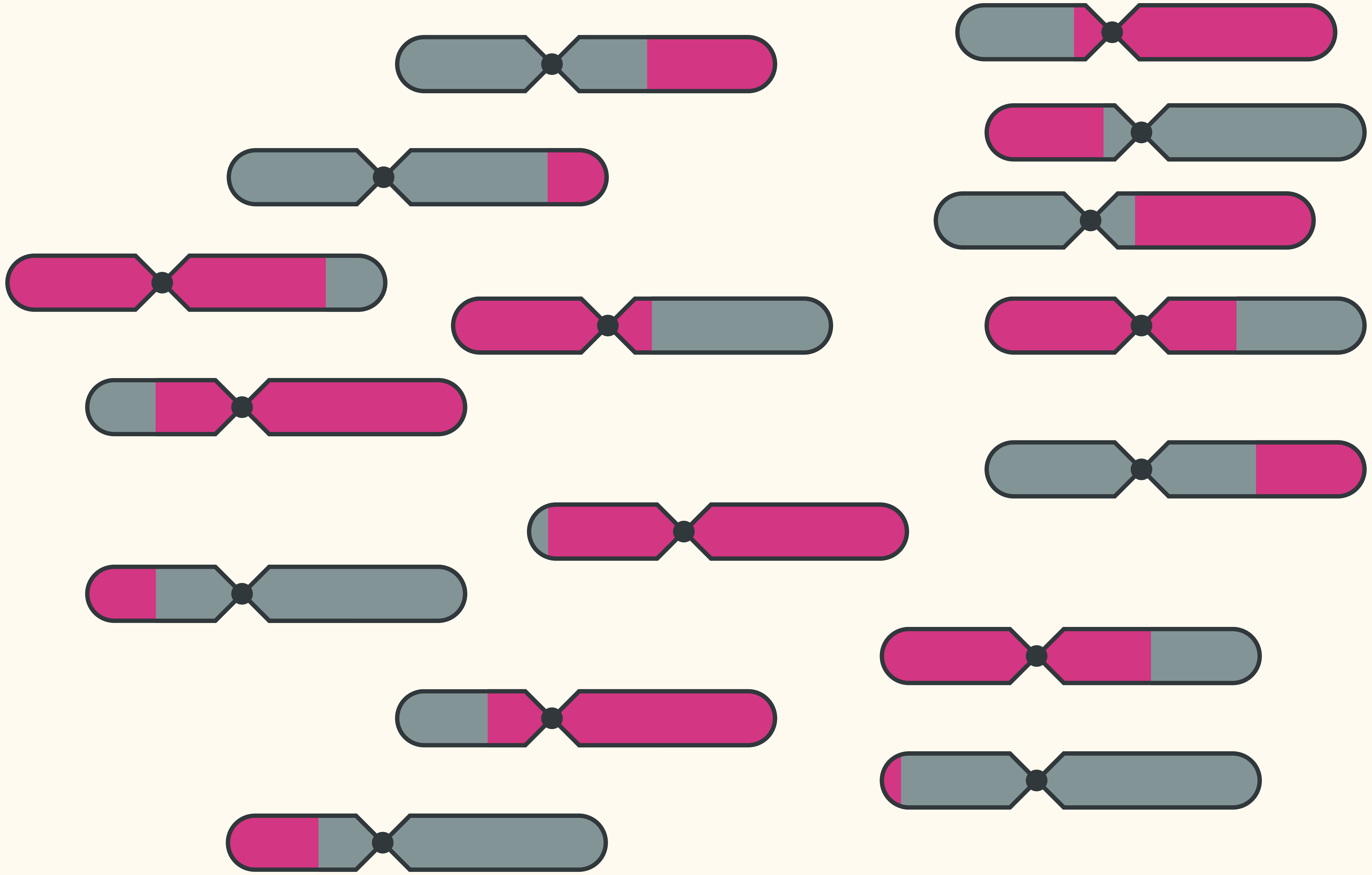


Hybridization

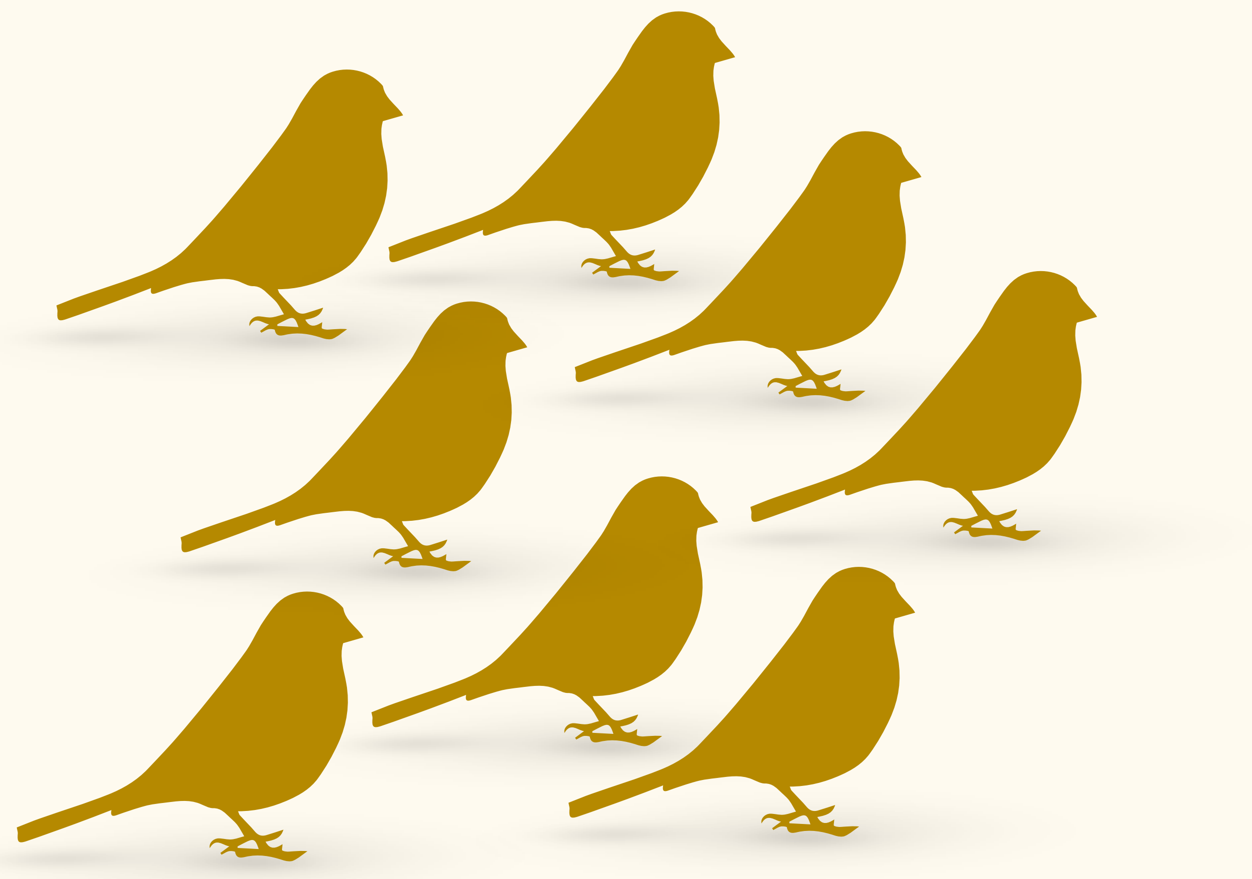


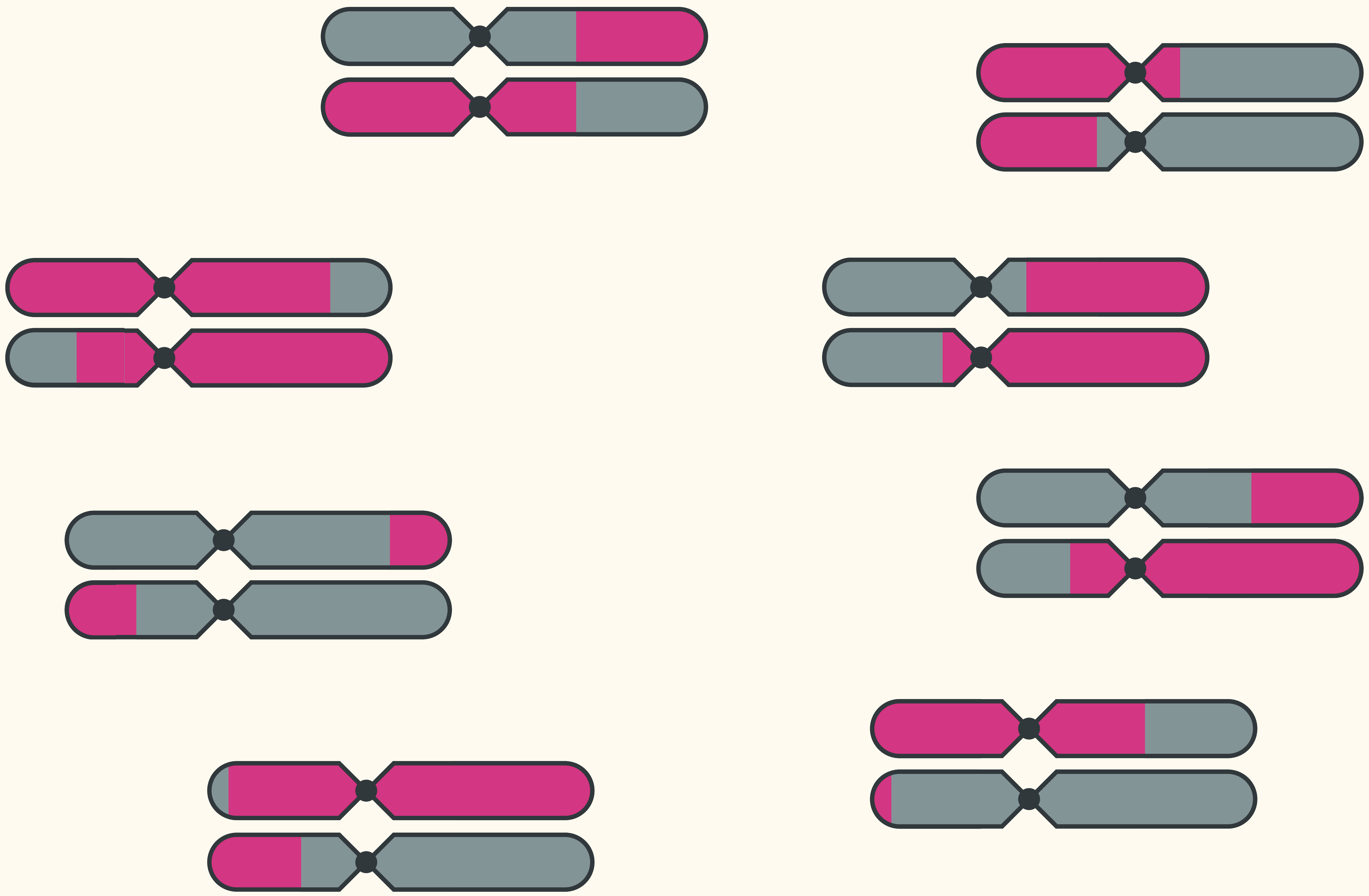
Hybridization



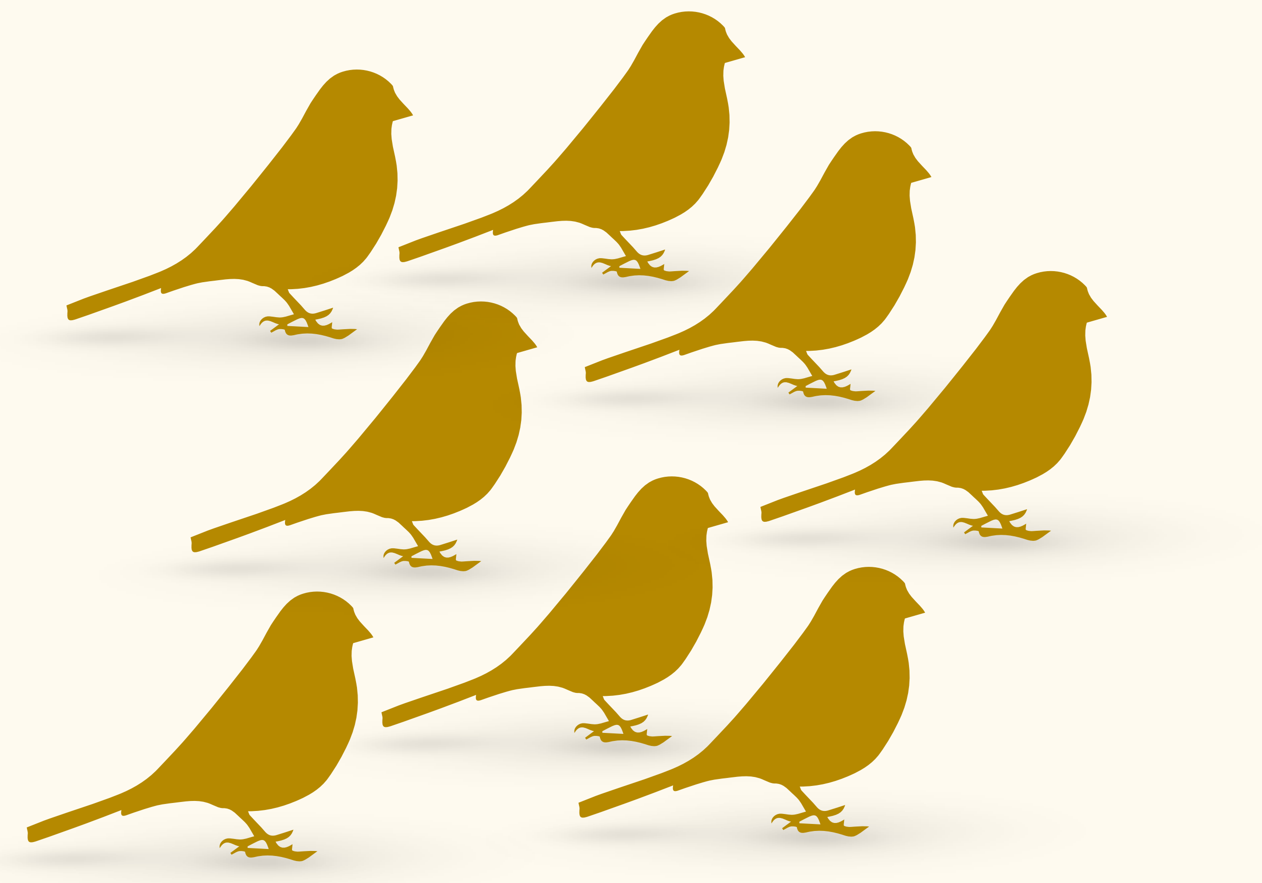


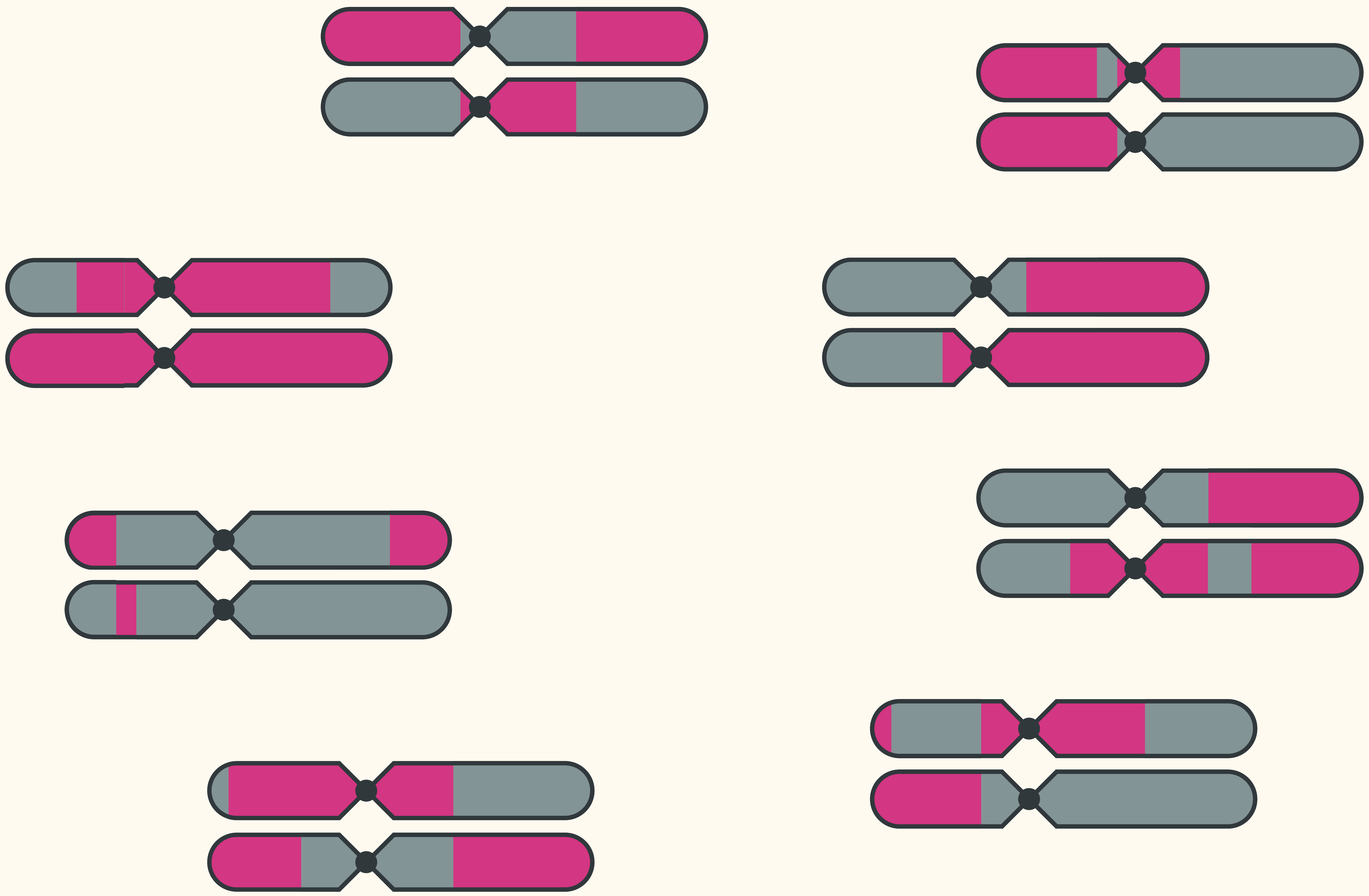
Hybridization



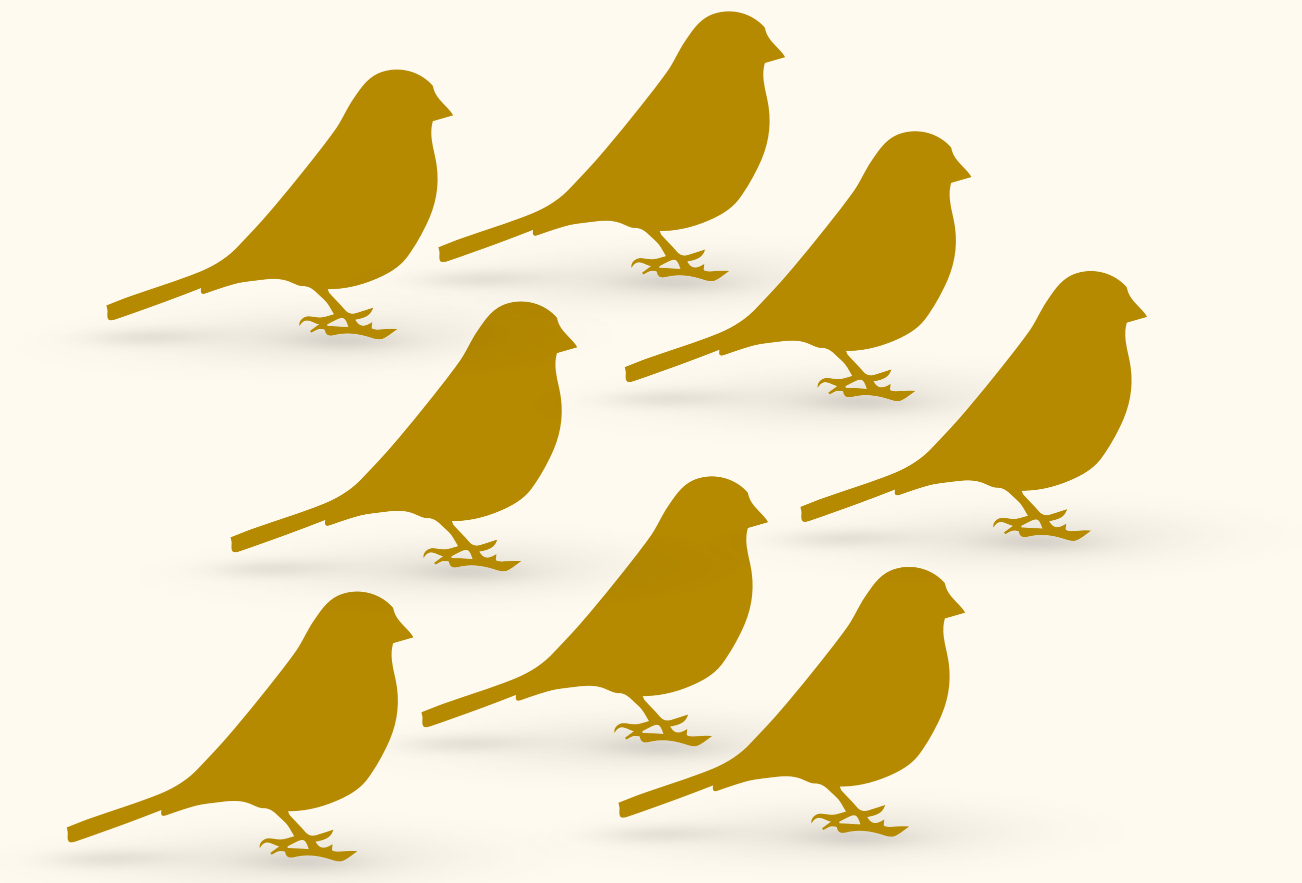


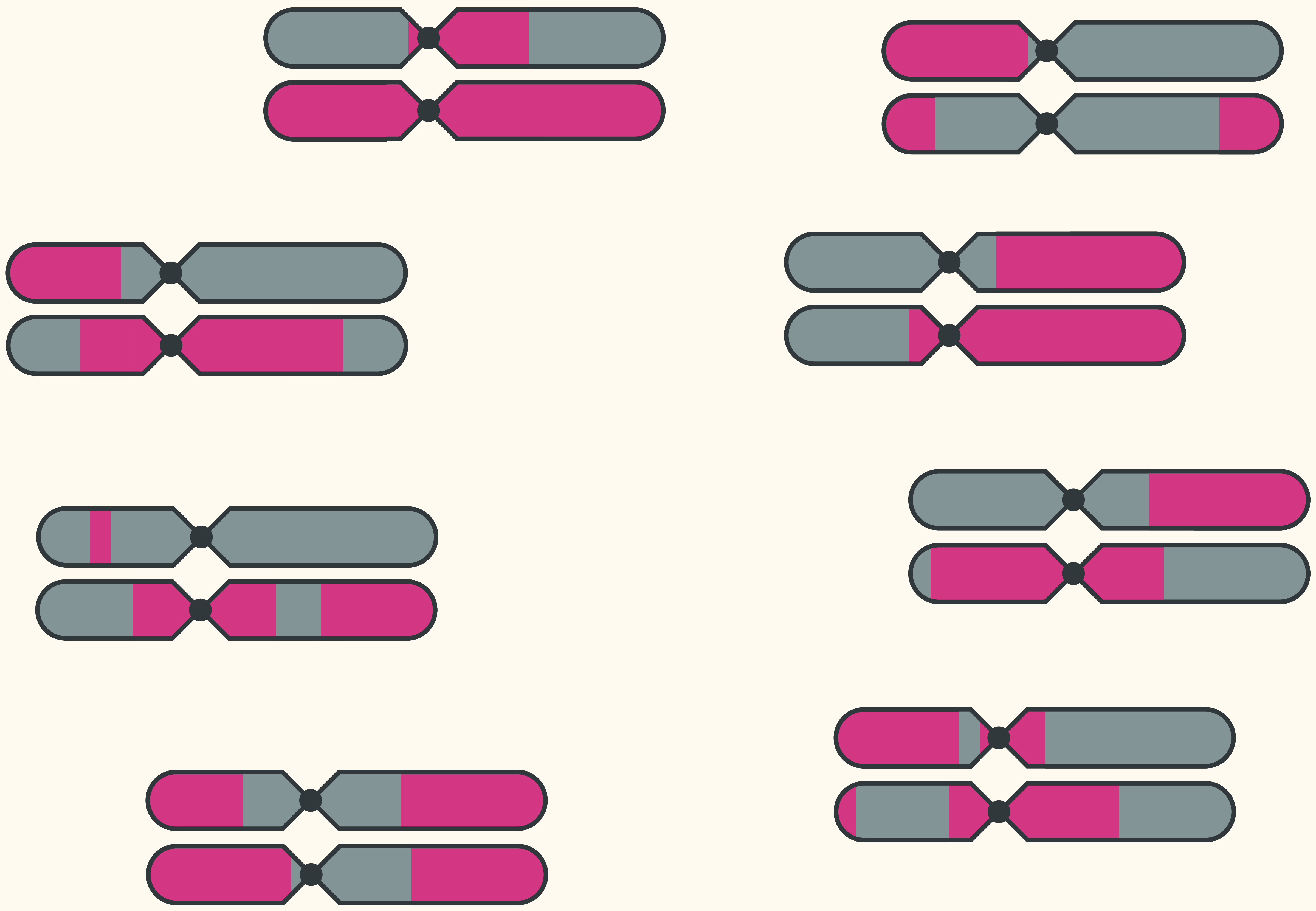
Hybridization



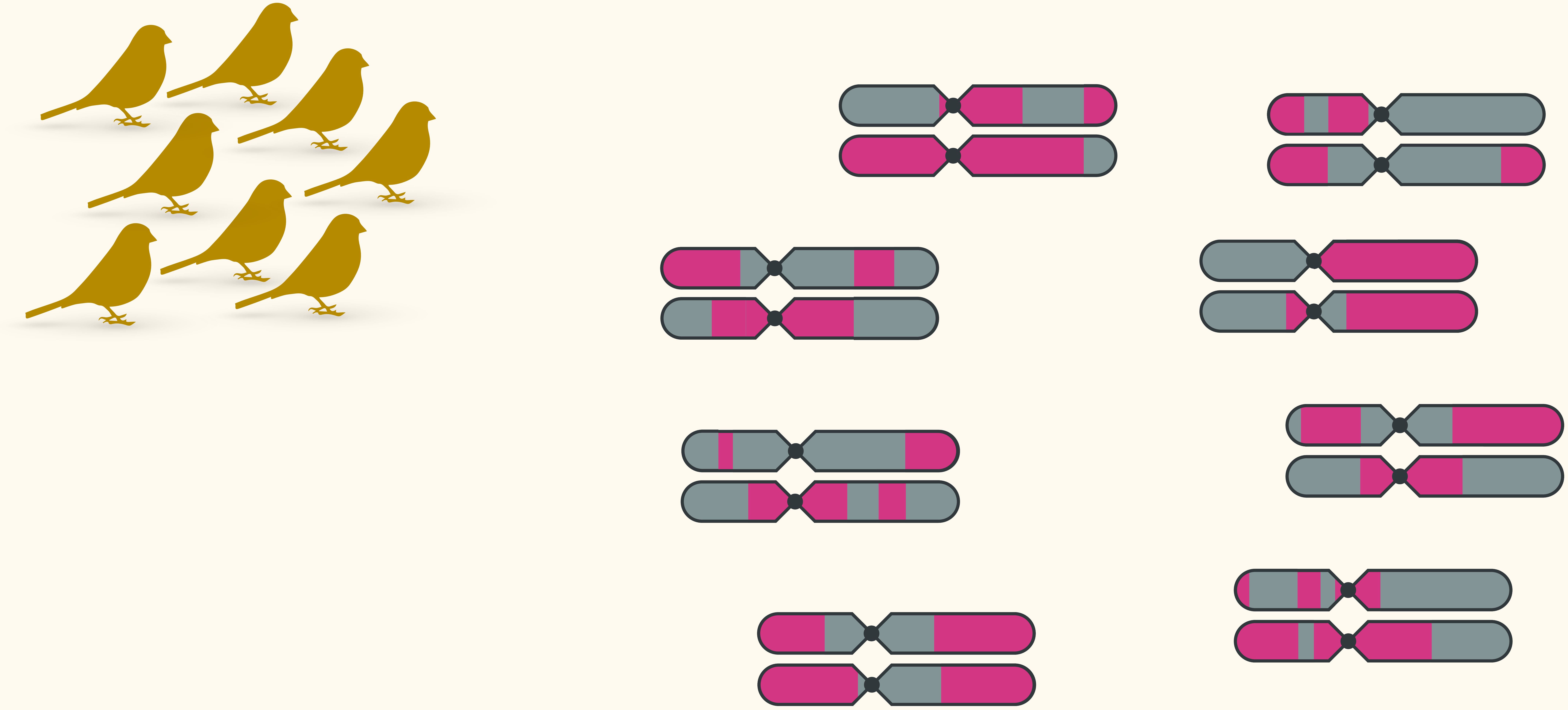


Hybridization

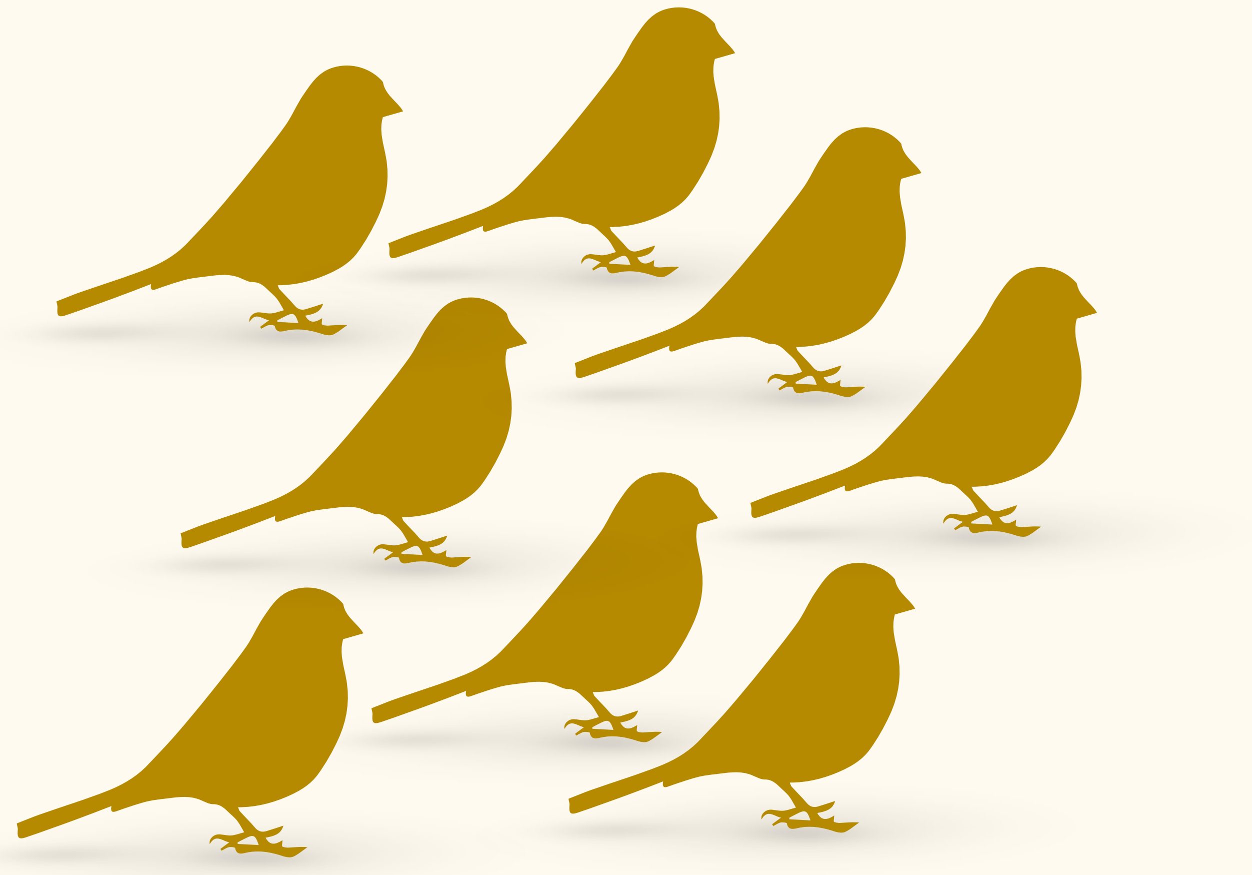


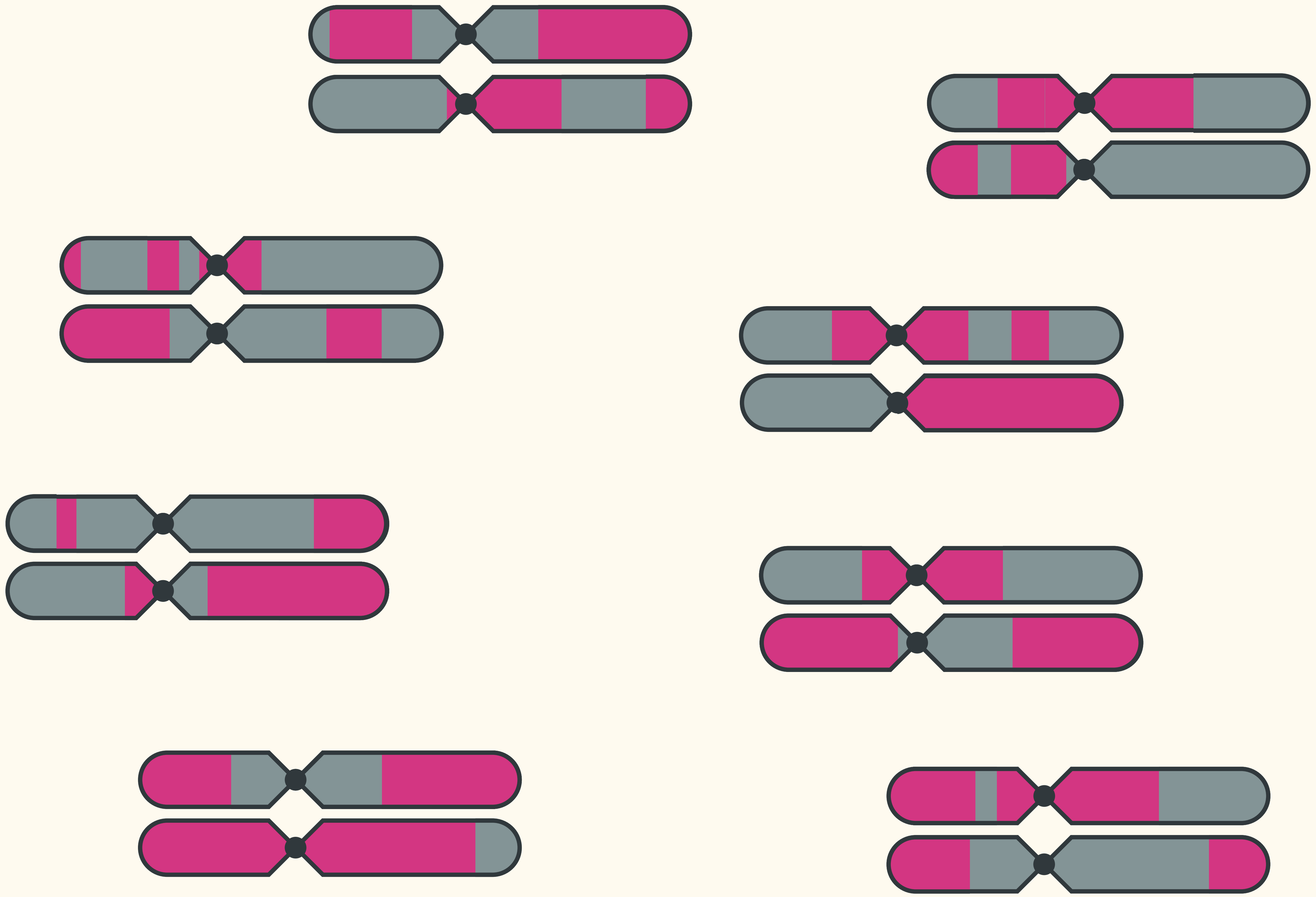


Hybridization

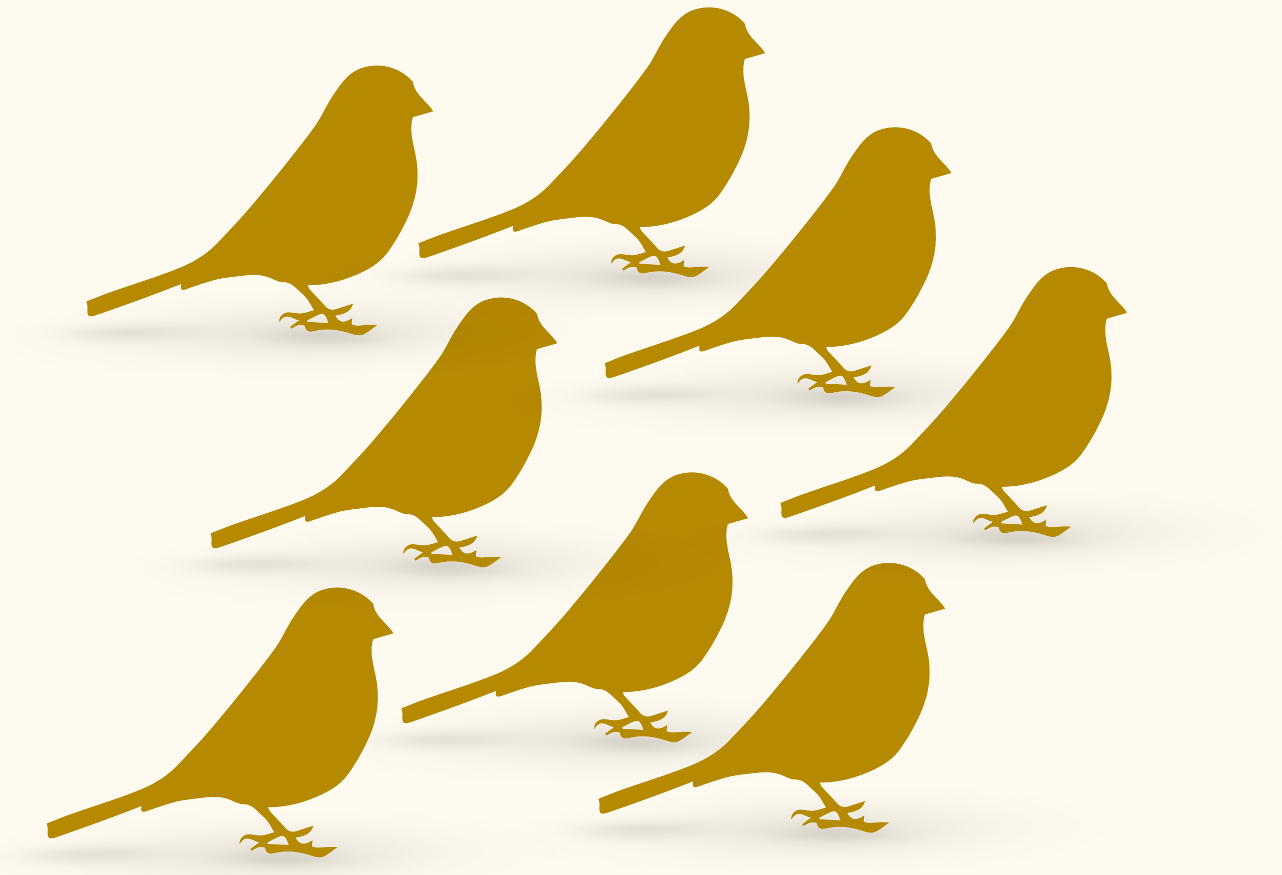


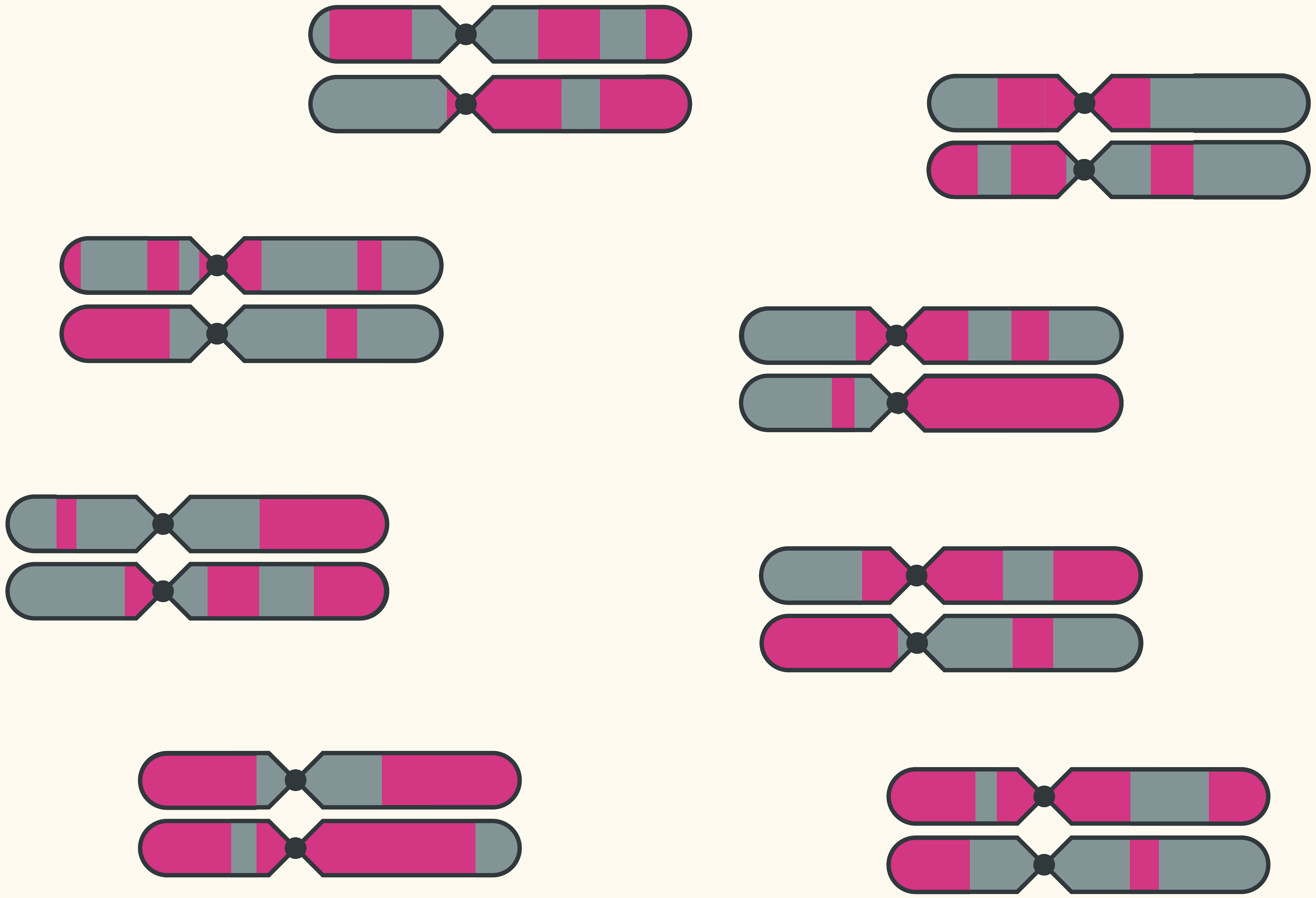
Hybridization





Hybridization





Hybridization

Photos: Francesco Veronesi / Wikipedia

Hybridization

Italian Sparrow

Photos: Francesco Veronesi / Wikipedia

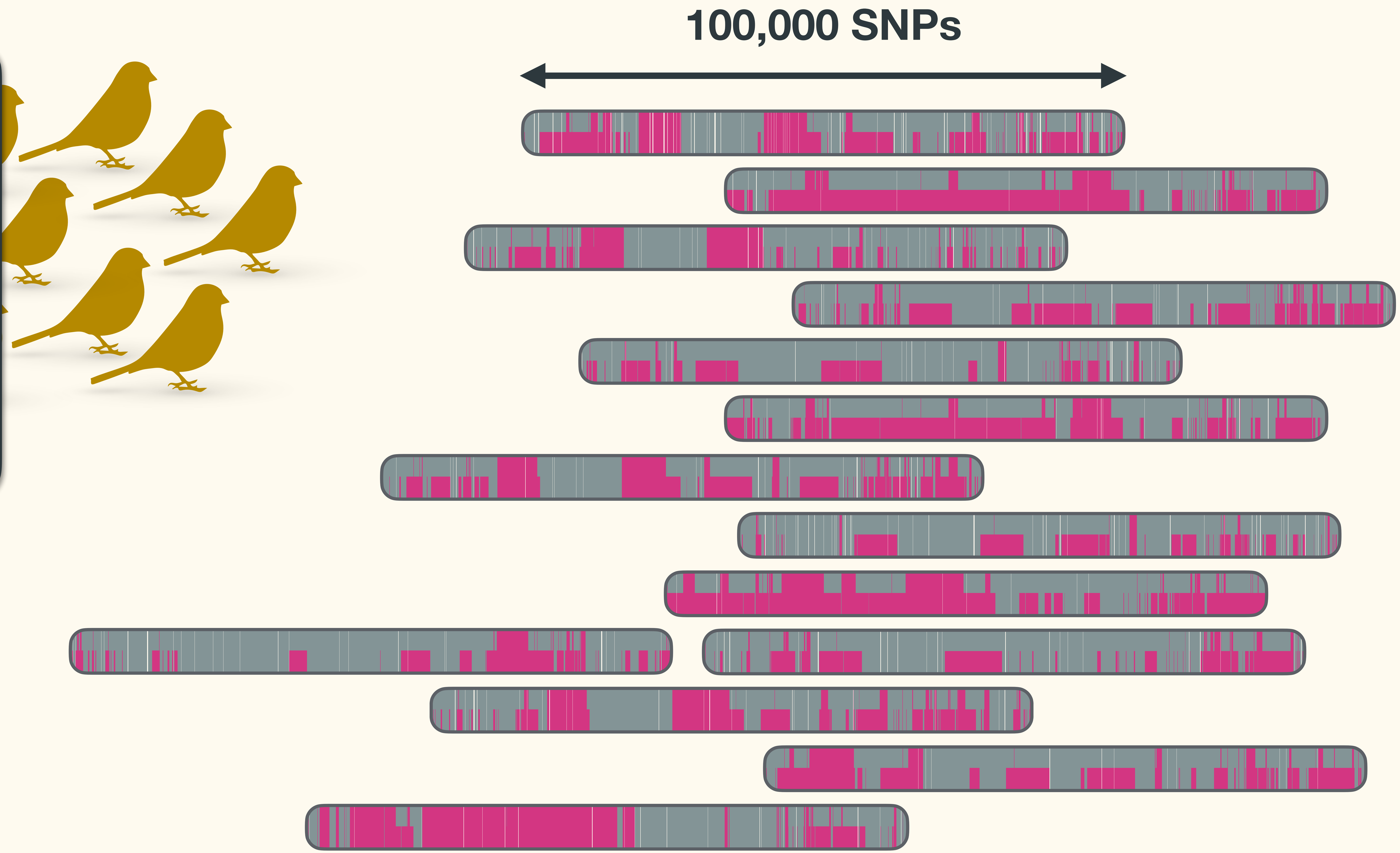
Hybridization

Italian Sparrow
House Sparrow
Spanish Sparrow

Photos: Dgwildlife / Getty; Francesco Veronesi / Wikipedia // Ref.: Runemark et al. (2018) *Nature Ecology and Evolution* 2:549–556

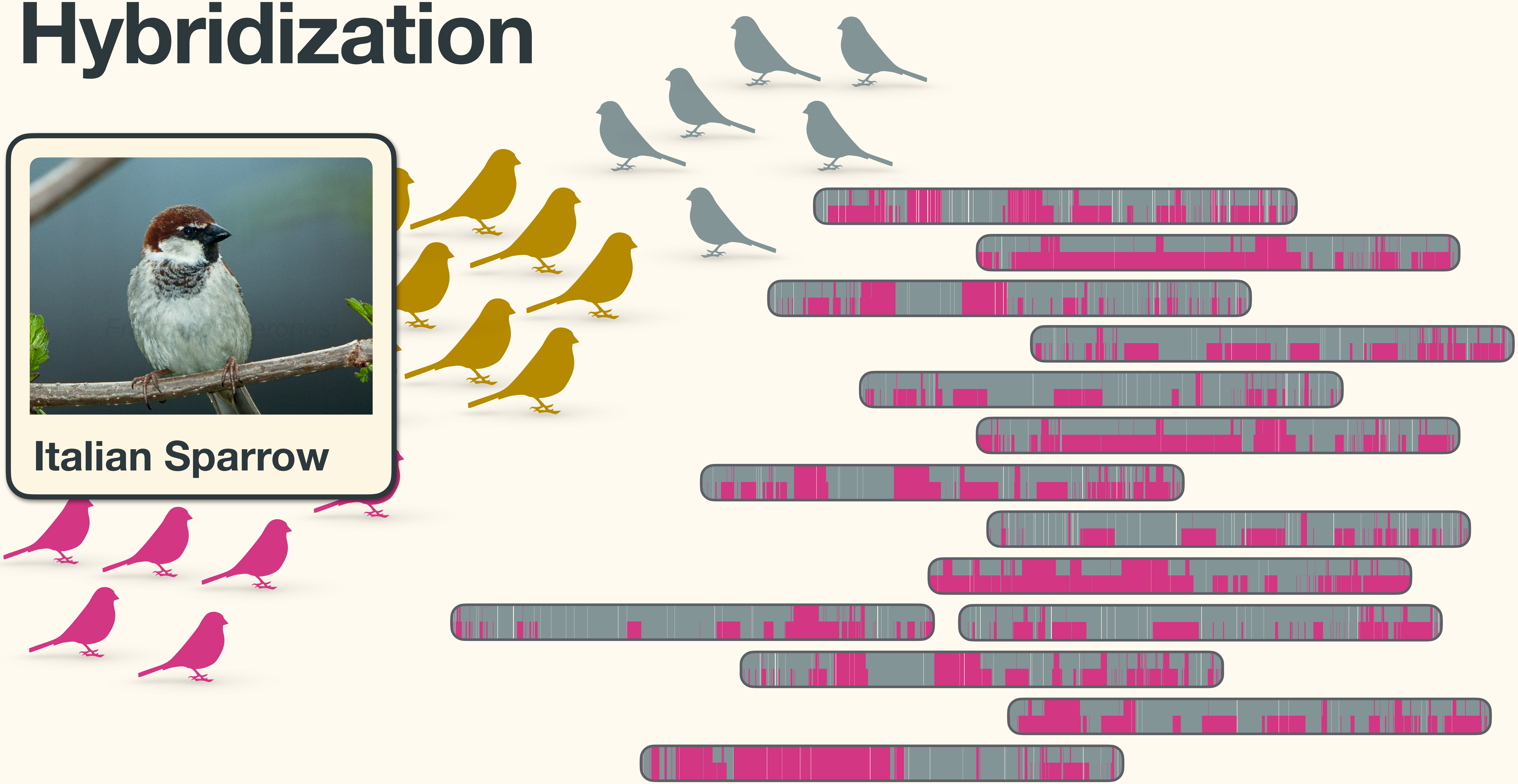
Hybridization

Italian Sparrow



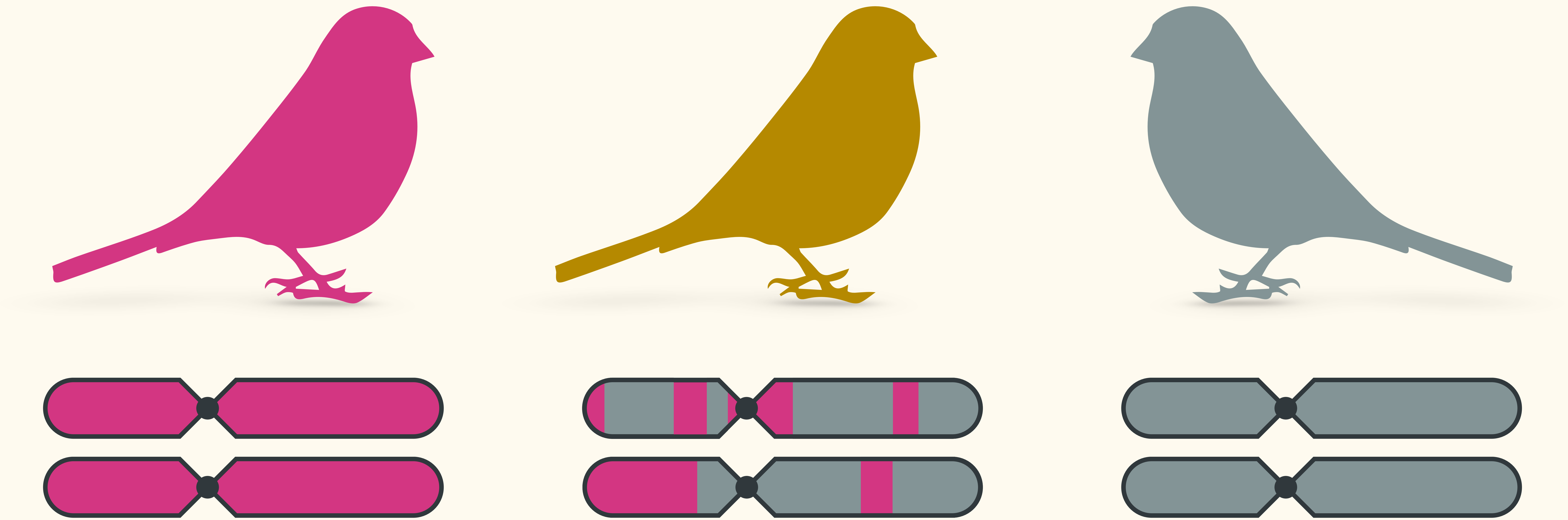
Ref.: Runemark et al. (2018) *Nature Ecology and Evolution* 2:549–556

Hybridization

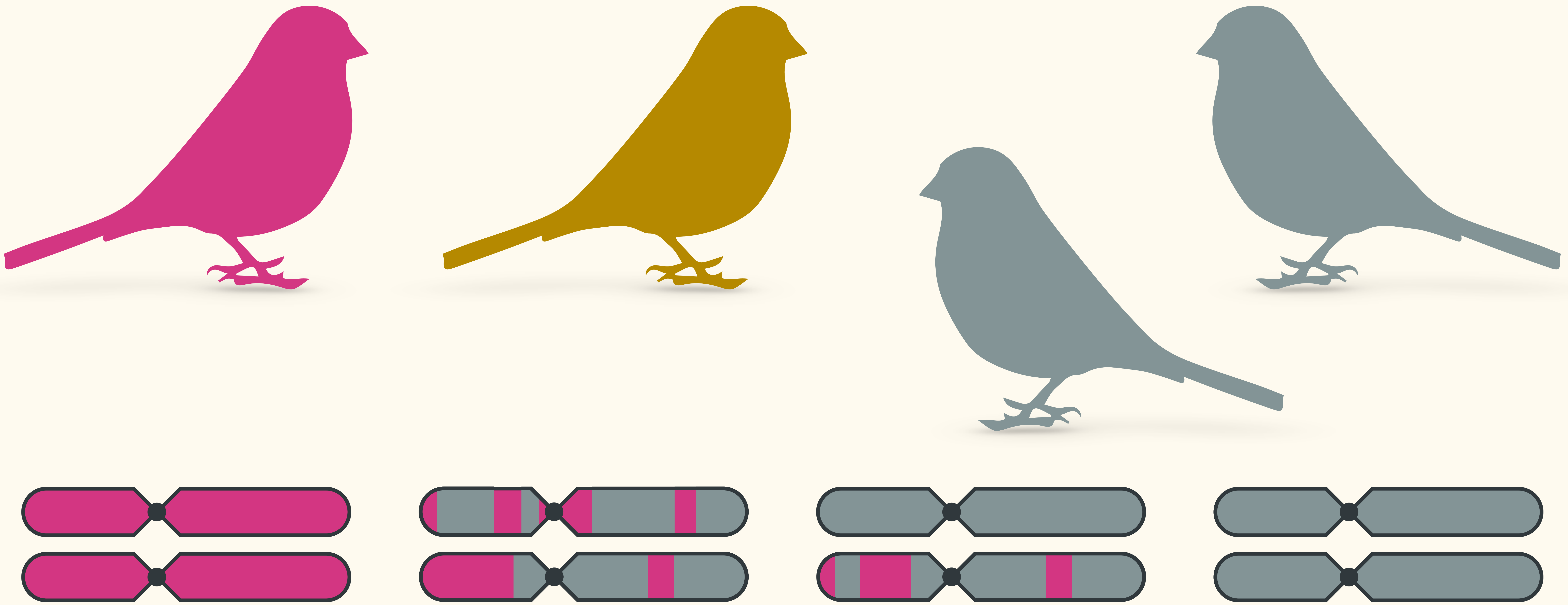


Ref.: Runemark et al. (2018) *Nature Ecology and Evolution* 2:549–556

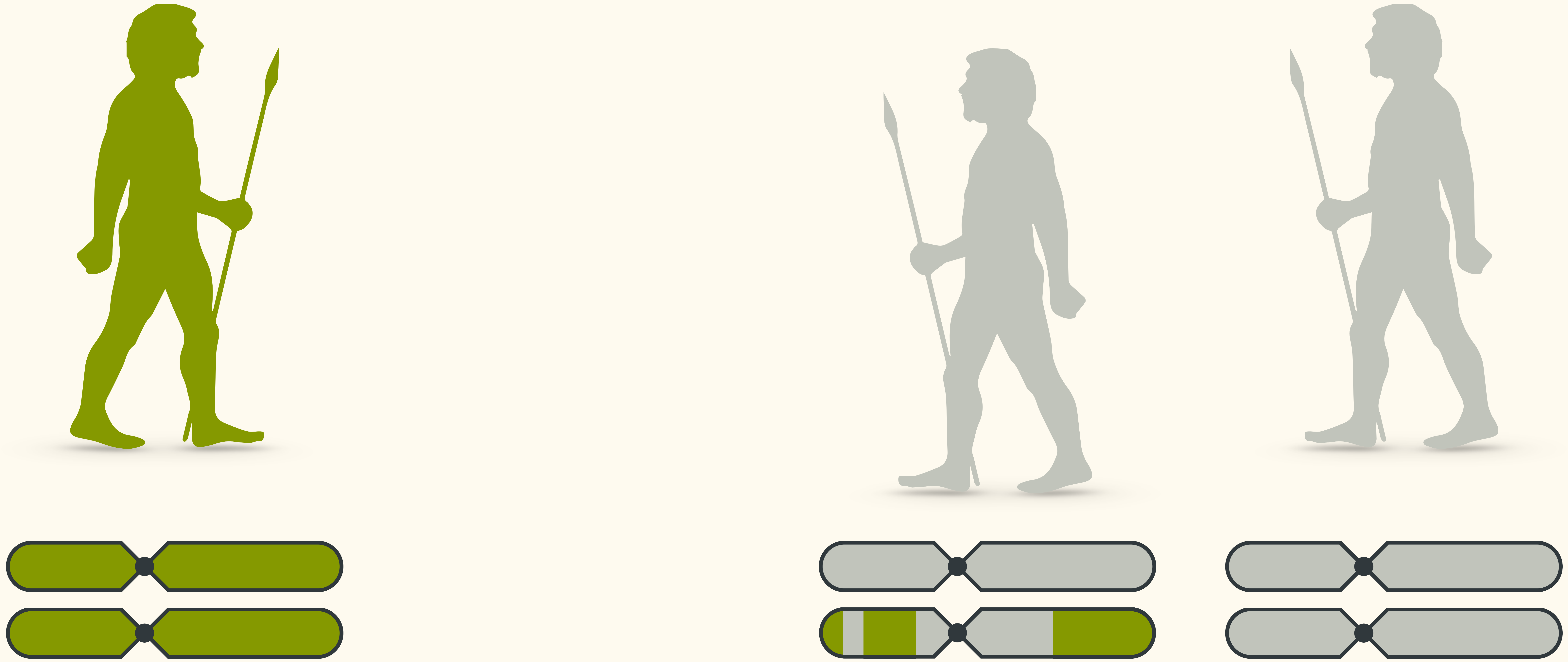
Introgression



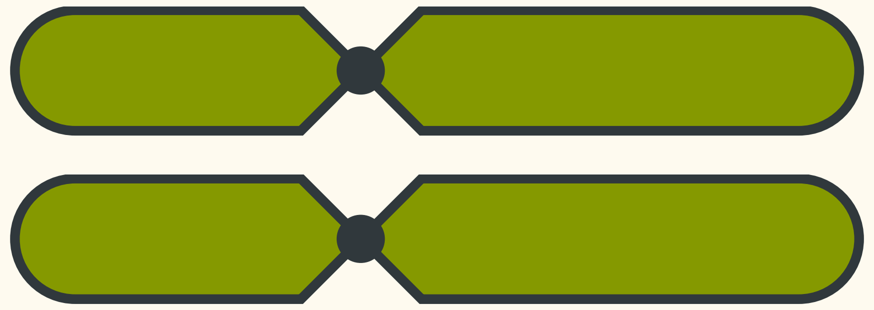
Introgression



Introgression



Introgression



LETTER

doi:10.1038/nature14558

An early modern human from Romania with a recent Neanderthal ancestor

Qiaomei Fu[1,2,3]*, Mateja Hajdinjak[3]*, Oana Teodora Moldovan[4], Silviu Constantin[5], Swapan Mallick[2,6,7], Pontus Skoglund[2], Nick Patterson[6], Nadin Rohland[2], Iosif Lazaridis[2], Birgit Nickel[3], Bence Viola[3,7,8], Kay Prüfer[3], Matthias Meyer[3], Janet Kelso[3], David Reich[2,6,9] & Svante Pääbo[3]

Neanderthals are thought to have disappeared in Europe approximately 39,000–41,000 years ago but they have contributed 1–3% of the DNA of present-day people in Eurasia[1]. Here we analyse DNA from a 37,000–42,000-year-old[2] modern human from Peştera cu Oase, Romania. Although the specimen contains small amounts of human DNA, we use an enrichment strategy to isolate sites that are informative about its relationship to Neanderthals and present-day humans. We find that on the order of 6–9% of the genome of the Oase individual is derived from Neanderthals, more than any other modern human sequenced to date. Three chromosomal segments of Neanderthal ancestry are over 50 centimorgans in size, indicating that this individual had a Neanderthal ancestor as recently as four to six generations back. However, the Oase individual does not share more alleles with later Europeans than with East Asians, suggesting that the Oase population did not contribute substantially to later humans in Europe.

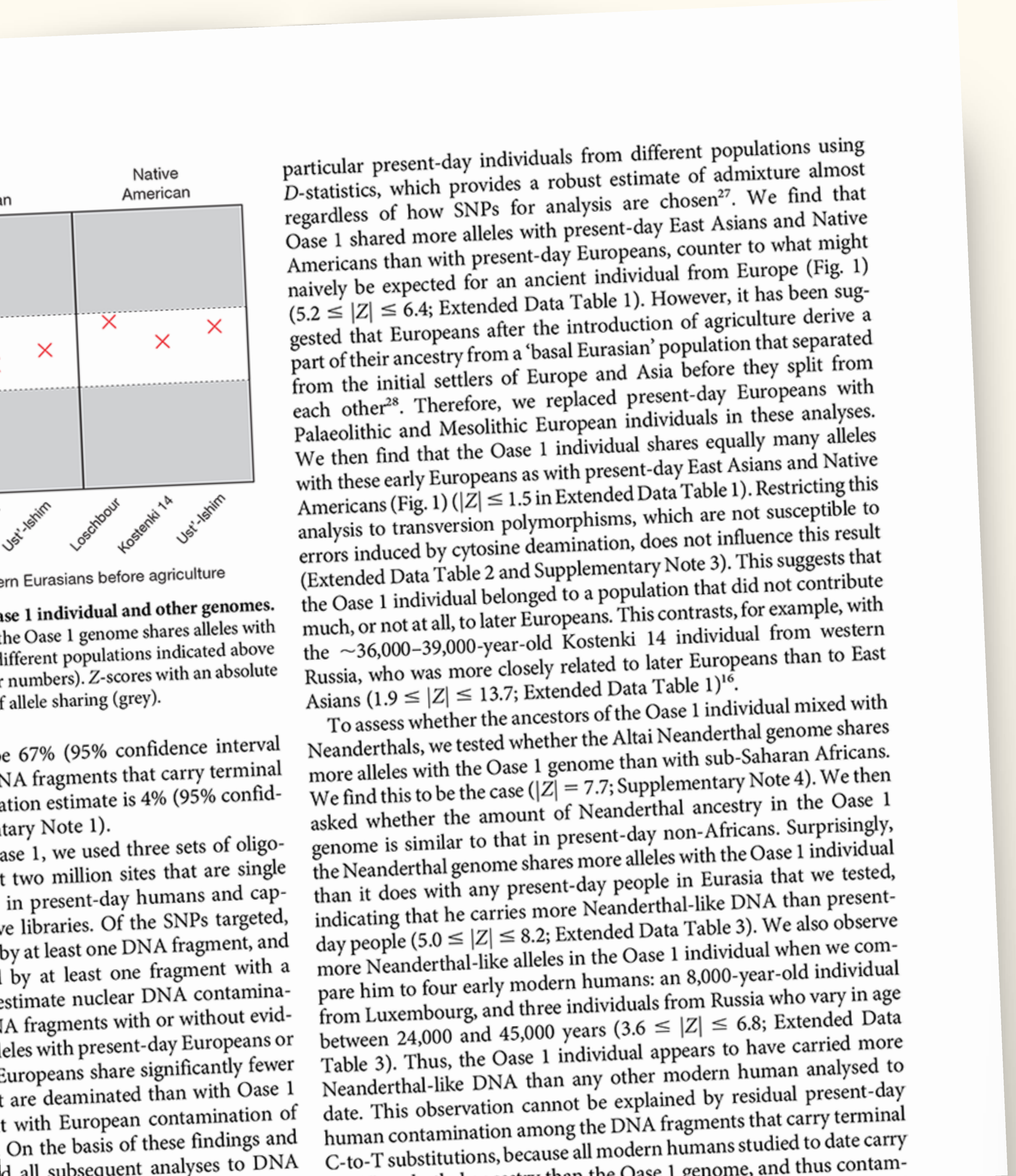
Between 45,000 and 35,000 years ago, anatomically modern humans spread across Europe, while the Neanderthals, present since before 300,000 years ago, disappeared. How this process occurred has long been debated[1,3–5]. Comparisons between the Neanderthal genome and the genomes of present-day humans have shown that Neanderthals contributed approximately 1–3% of the genomes of all people living today outside sub-Saharan Africa[6,7] suggesting that human populations ancestral to all non-Africans mixed with Neanderthals. The size of segments of Neanderthal ancestry in present-day humans suggests that this occurred between 37,000 and 86,000 years ago[8]. However, where and how often this occurred is not understood. For example, Neanderthals share

We report genome-wide data from a modern human mandible, Oase 1, found in 2002 in the Peştera cu Oase, Romania. The age of this specimen has been estimated to be ~37,000–42,000 years by direct radiocarbon dating[2,17,18]. Oase 1 is therefore one of the earliest modern humans in Europe. Its morphology is generally modern but some aspects are consistent with Neanderthal ancestry[19–21]. Subsequent excavations uncovered a cranium from another, probably contemporaneous individual, Oase 2, which also carries morphological traits that could reflect admixture with Neanderthals[17,19].

We prepared two DNA extracts from 25 mg and 10 mg of bone powder removed from the inferior right ramus of Oase 1. We treated an aliquot of each of these extracts with *Escherichia coli* uracil-DNA glycosylase (UDG), an enzyme that removes uracils from the interior parts of DNA molecules, but leaves a proportion of uracils at the ends of the molecules unaffected. Uracil residues occur in DNA molecules as a result of deamination of cytosine residues, and are particularly prevalent at the ends of ancient DNA molecules[9,22]. Among the DNA fragments sequenced from these two extracts, 0.18% and 0.06%, respectively, could be mapped to the human reference genome. We prepared three additional DNA libraries from the extract containing 0.18% human-like molecules, but omitted the UDG treatment to increase the number of molecules in which terminal C-to-T substitutions could be seen and used to identify putatively ancient fragments. Because the fraction of endogenous DNA is so small, we used hybridization to DNA probes to isolate human DNA fragments from the libraries[23]. Applying this strategy to the mitochondrial genome allowed the mitochondrial (mt)DNA

524:216-219

Introgression



ase 1 individual and other genomes. he Oase 1 genome shares alleles with lifferent populations indicated above r numbers). Z-scores with an absolute f allele sharing (grey).

e 67% (95% confidence interval NA fragments that carry terminal ation estimate is 4% (95% confid- tary Note 1).

ase 1, we used three sets of oligo- t two million sites that are single in present-day humans and cap- ve libraries. Of the SNPs targeted, by at least one DNA fragment, and l by at least one fragment with a estimate nuclear DNA contamina- IA fragments with or without evid- leles with present-day Europeans or Europeans share significantly fewer t are deaminated than with Oase 1 t with European contamination of On the basis of these findings and d all subsequent analyses to DNA

particular present-day individuals from different populations using *D*-statistics, which provides a robust estimate of admixture almost regardless of how SNPs for analysis are chosen[27]. We find that Oase 1 shared more alleles with present-day East Asians and Native Americans than with present-day Europeans, counter to what might naively be expected for an ancient individual from Europe (Fig. 1) ($5.2 \le |Z| \le 6.4$; Extended Data Table 1). However, it has been suggested that Europeans after the introduction of agriculture derive a part of their ancestry from a 'basal Eurasian' population that separated from the initial settlers of Europe and Asia before they split from each other[28]. Therefore, we replaced present-day Europeans with Palaeolithic and Mesolithic European individuals in these analyses. We then find that the Oase 1 individual shares equally many alleles with these early Europeans as with present-day East Asians and Native Americans (Fig. 1) ($|Z| \le 1.5$ in Extended Data Table 1). Restricting this analysis to transversion polymorphisms, which are not susceptible to errors induced by cytosine deamination, does not influence this result (Extended Data Table 2 and Supplementary Note 3). This suggests that the Oase 1 individual belonged to a population that did not contribute much, or not at all, to later Europeans. This contrasts, for example, with the ~36,000–39,000-year-old Kostenki 14 individual from western Russia, who was more closely related to later Europeans than to East Asians ($1.9 \le |Z| \le 13.7$; Extended Data Table 1)[16].

To assess whether the ancestors of the Oase 1 individual mixed with Neanderthals, we tested whether the Altai Neanderthal genome shares more alleles with the Oase 1 genome than with sub-Saharan Africans. We find this to be the case ($|Z| = 7.7$; Supplementary Note 4). We then asked whether the amount of Neanderthal ancestry in the Oase 1 genome is similar to that in present-day non-Africans. Surprisingly, the Neanderthal genome shares more alleles with the Oase 1 individual than it does with any present-day people in Eurasia that we tested, indicating that he carries more Neanderthal-like DNA than present-day people ($5.0 \le |Z| \le 8.2$; Extended Data Table 3). We also observe more Neanderthal-like alleles in the Oase 1 individual when we compare him to four early modern humans: an 8,000-year-old individual from Luxembourg, and three individuals from Russia who vary in age between 24,000 and 45,000 years ($3.6 \le |Z| \le 6.8$; Extended Data Table 3). Thus, the Oase 1 individual appears to have carried more Neanderthal-like DNA than any other modern human analysed to date. This observation cannot be explained by residual present-day human contamination among the DNA fragments that carry terminal C-to-T substitutions, because all modern humans studied to date carry

LETTER

doi:10.1038/nature14558

An early modern human from Romania with a recent Neanderthal ancestor

Qiaomei Fu[1,2,3]*, Mateja Hajdinjak[3]*, Oana Teodora Moldovan[4], Silviu Constantin[5], Swapan Mallick[2,6,7], Pontus Skoglund[2], Nick Patterson[6], Nadin Rohland[2], Iosif Lazaridis[2], Birgit Nickel[3], Bence Viola[3,7,8], Kay Prüfer[3], Matthias Meyer[3], Janet Kelso[3], David Reich[2,6,9] & Svante Pääbo[3]

Neanderthals are thought to have disappeared in Europe approximately 39,000–41,000 years ago but they have contributed 1–3% of the DNA of present-day people in Eurasia[1]. Here we analyse DNA from a 37,000–42,000-year-old[2] modern human from Peştera cu Oase, Romania. Although the specimen contains small amounts of human DNA, we use an enrichment strategy to isolate sites that are informative about its relationship to Neanderthals and present-day humans. We find that on the order of 6–9% of the genome of the Oase individual is derived from Neanderthals, more than any other modern human sequenced to date. Three chromosomal segments of Neanderthal ancestry are over 50 centimorgans in size, indicating that this individual had a Neanderthal ancestor as recently as four to six generations back. However, the Oase individual does not share more alleles with later Europeans than with East Asians, suggesting that the Oase population did not contribute substantially to later humans in Europe.

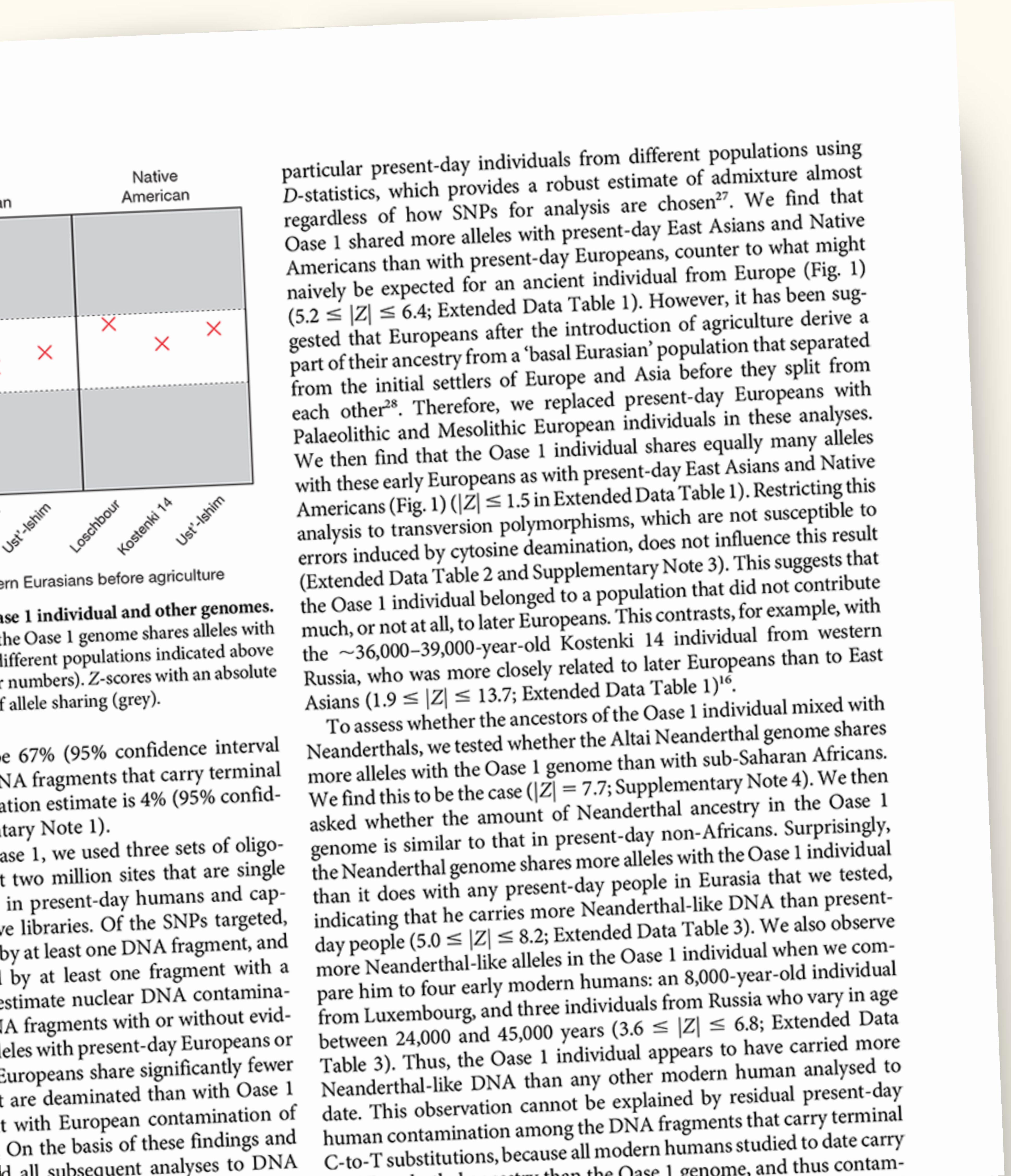
Between 45,000 and 35,000 years ago, anatomically modern humans spread across Europe, while the Neanderthals, present since before 300,000 years ago, disappeared. How this process occurred has long been debated[1,3–5]. Comparisons between the Neanderthal genome and the genomes of present-day humans have shown that Neanderthals contributed approximately 1–3% of the genomes of all people living today outside sub-Saharan Africa[6,7] suggesting that human populations ancestral to all non-Africans mixed with Neanderthals. The size of segments of Neanderthal ancestry in present-day humans suggests that this occurred between 37,000 and 86,000 years ago[8]. However, where and how often this occurred is not understood. For example, Neanderthals share more alleles

We report genome-wide data from a modern human mandible, Oase 1, found in 2002 in the Peştera cu Oase, Romania. The age of this specimen has been estimated to be ~37,000–42,000 years by direct radiocarbon dating[2,17,18]. Oase 1 is therefore one of the earliest modern humans in Europe. Its morphology is generally modern but some aspects are consistent with Neanderthal ancestry[19–21]. Subsequent excavations uncovered a cranium from another, probably contemporaneous individual, Oase 2, which also carries morphological traits that could reflect admixture with Neanderthals[17,19].

We prepared two DNA extracts from 25 mg and 10 mg of bone powder removed from the inferior right ramus of Oase 1. We treated an aliquot of each of these extracts with *Escherichia coli* uracil-DNA glycosylase (UDG), an enzyme that removes uracils from the interior parts of DNA molecules, but leaves a proportion of uracils at the ends of the molecules unaffected. Uracil residues occur in DNA molecules as a result of deamination of cytosine residues, and are particularly prevalent at the ends of ancient DNA molecules[9,22]. Among the DNA fragments sequenced from these two extracts, 0.18% and 0.06%, respectively, could be mapped to the human reference genome. We prepared three additional DNA libraries from the extract containing 0.18% human-like molecules, but omitted the UDG treatment to increase the number of molecules in which terminal C-to-T substitutions could be seen and used to identify putatively ancient fragments. Because the fraction of endogenous DNA is so small, we used hybridization to DNA probes to isolate human DNA fragments from the libraries[23]. Applying this strategy to the mitochondrial genome allowed the mitochondrial (mt)DNA from the

524:216-219

Introgression



Native American
Ust'-Ishim
Loschbour
Kostenki 14
Ust'-Ishim
ern Eurasians before agriculture

ase 1 individual and other genomes. the Oase 1 genome shares alleles with different populations indicated above r numbers). Z-scores with an absolute f allele sharing (grey).

e 67% (95% confidence interval NA fragments that carry terminal ation estimate is 4% (95% confid- tary Note 1).

ase 1, we used three sets of oligo- t two million sites that are single in present-day humans and cap- ve libraries. Of the SNPs targeted, by at least one DNA fragment, and l by at least one fragment with a estimate nuclear DNA contamina- IA fragments with or without evid- leles with present-day Europeans or Europeans share significantly fewer t are deaminated than with Oase 1 t with European contamination of On the basis of these findings and d all subsequent analyses to DNA

particular present-day individuals from different populations using *D*-statistics, which provides a robust estimate of admixture almost regardless of how SNPs for analysis are chosen[27]. We find that Oase 1 shared more alleles with present-day East Asians and Native Americans than with present-day Europeans, counter to what might naively be expected for an ancient individual from Europe (Fig. 1) (5.2 ≤ |Z| ≤ 6.4; Extended Data Table 1). However, it has been suggested that Europeans after the introduction of agriculture derive a part of their ancestry from a 'basal Eurasian' population that separated from the initial settlers of Europe and Asia before they split from each other[28]. Therefore, we replaced present-day Europeans with Palaeolithic and Mesolithic European individuals in these analyses. We then find that the Oase 1 individual shares equally many alleles with these early Europeans as with present-day East Asians and Native Americans (Fig. 1) (|Z| ≤ 1.5 in Extended Data Table 1). Restricting this analysis to transversion polymorphisms, which are not susceptible to errors induced by cytosine deamination, does not influence this result (Extended Data Table 2 and Supplementary Note 3). This suggests that the Oase 1 individual belonged to a population that did not contribute much, or not at all, to later Europeans. This contrasts, for example, with the ~36,000–39,000-year-old Kostenki 14 individual from western Russia, who was more closely related to later Europeans than to East Asians (1.9 ≤ |Z| ≤ 13.7; Extended Data Table 1)[16].

To assess whether the ancestors of the Oase 1 individual mixed with Neanderthals, we tested whether the Altai Neanderthal genome shares more alleles with the Oase 1 genome than with sub-Saharan Africans. We find this to be the case (|Z| = 7.7; Supplementary Note 4). We then asked whether the amount of Neanderthal ancestry in the Oase 1 genome is similar to that in present-day non-Africans. Surprisingly, the Neanderthal genome shares more alleles with the Oase 1 individual than it does with any present-day people in Eurasia that we tested, indicating that he carries more Neanderthal-like DNA than present-day people (5.0 ≤ |Z| ≤ 8.2; Extended Data Table 3). We also observe more Neanderthal-like alleles in the Oase 1 individual when we compare him to four early modern humans: an 8,000-year-old individual from Luxembourg, and three individuals from Russia who vary in age between 24,000 and 45,000 years (3.6 ≤ |Z| ≤ 6.8; Extended Data Table 3). Thus, the Oase 1 individual appears to have carried more Neanderthal-like DNA than any other modern human analysed to date. This observation cannot be explained by residual present-day human contamination among the DNA fragments that carry terminal C-to-T substitutions, because all modern humans studied to date carry

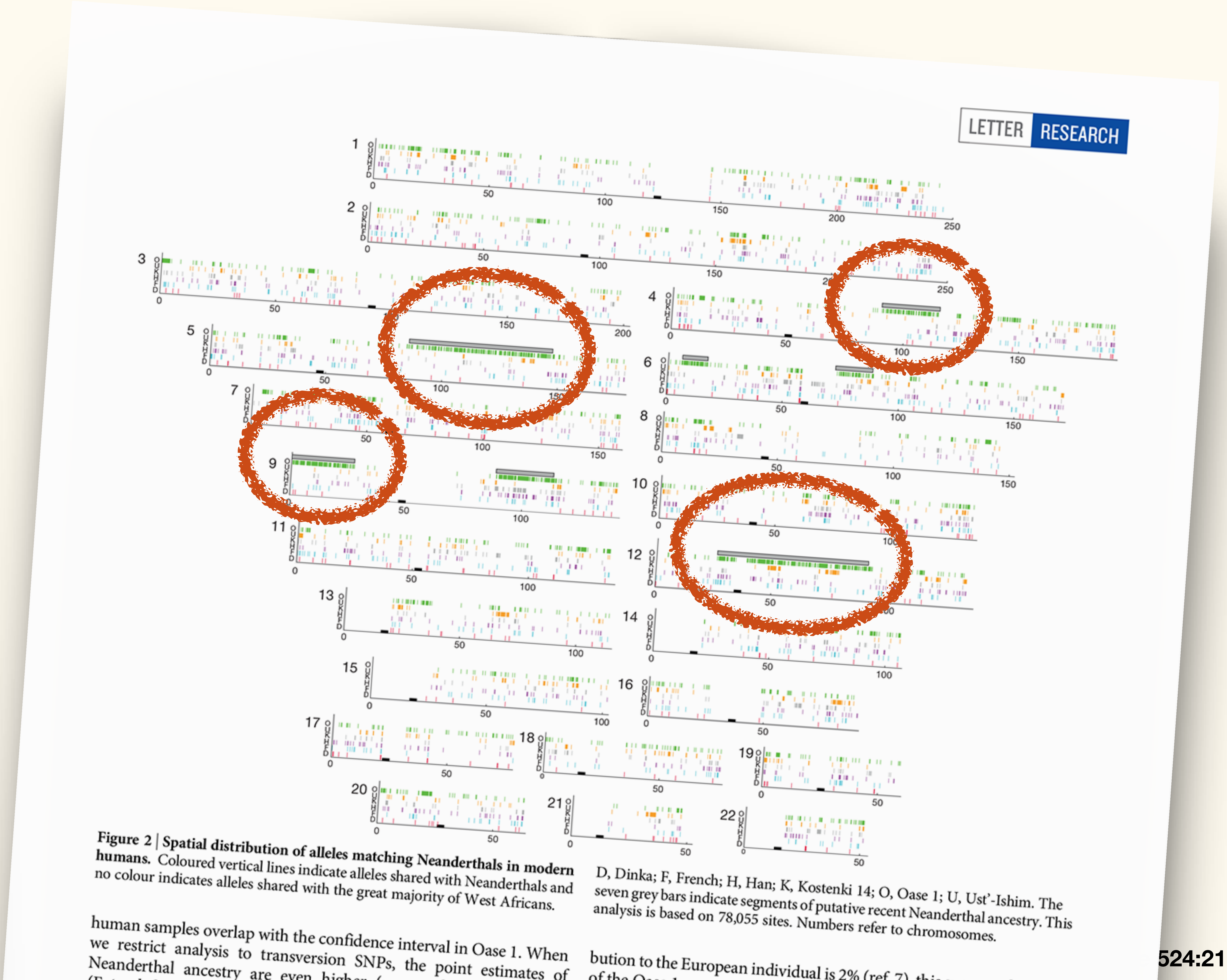

LETTER RESEARCH

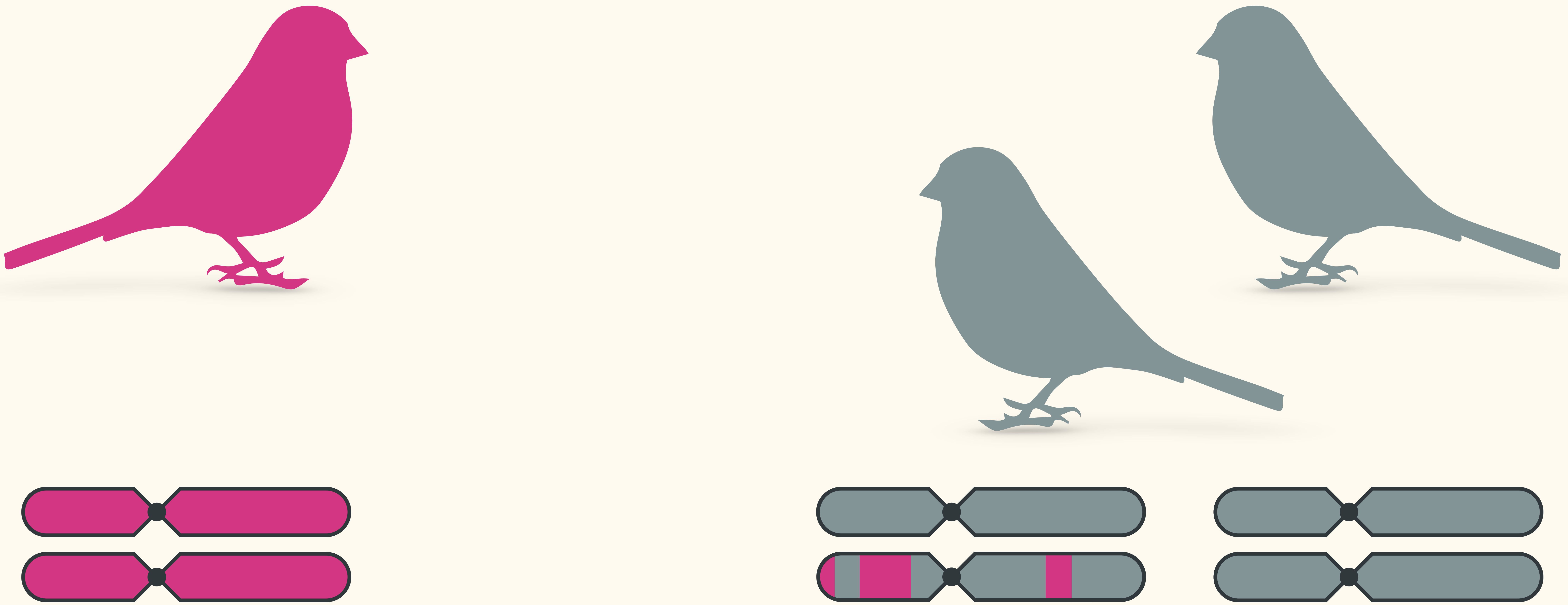
Figure 2 | Spatial distribution of alleles matching Neanderthals in modern humans. Coloured vertical lines indicate alleles shared with Neanderthals and no colour indicates alleles shared with the great majority of West Africans. D, Dinka; F, French; H, Han; K, Kostenki 14; O, Oase 1; U, Ust'-Ishim. The seven grey bars indicate segments of putative recent Neanderthal ancestry. This analysis is based on 78,055 sites. Numbers refer to chromosomes.

human samples overlap with the confidence interval in Oase 1. When we restrict analysis to transversion SNPs, the point estimates of Neanderthal ancestry are even higher

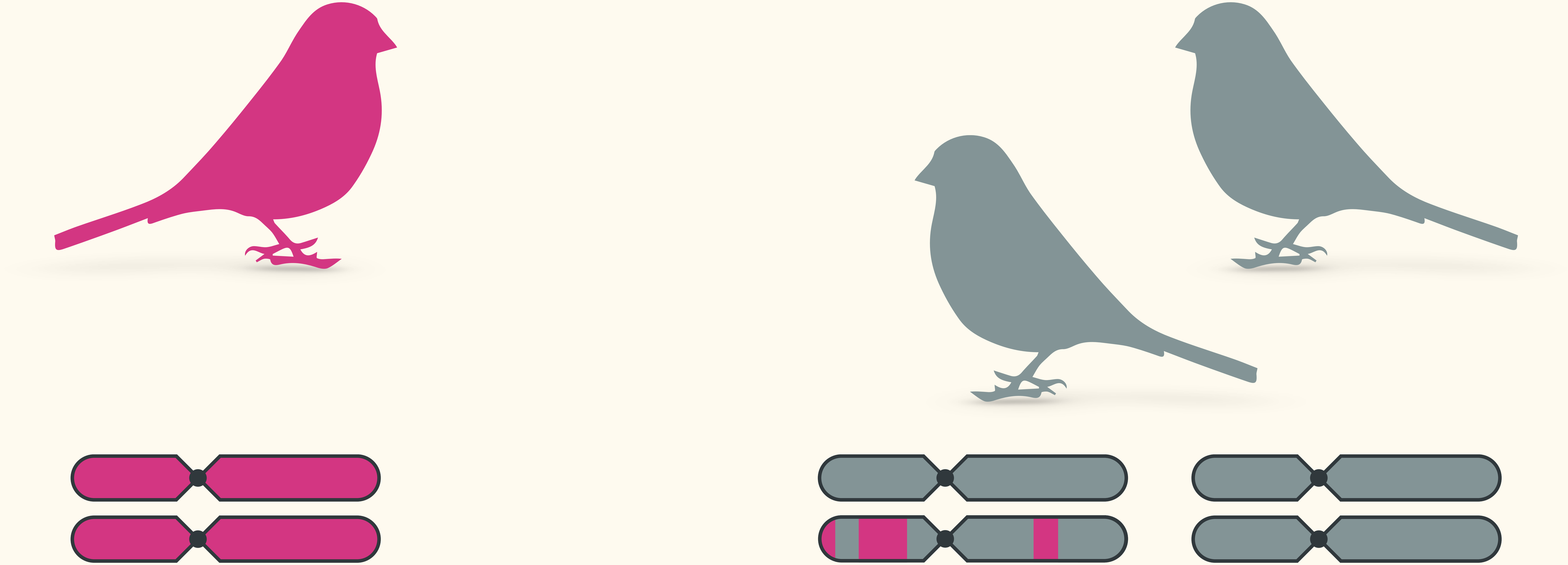
bution to the European individual is 2% (ref. 7), this

524:216-219

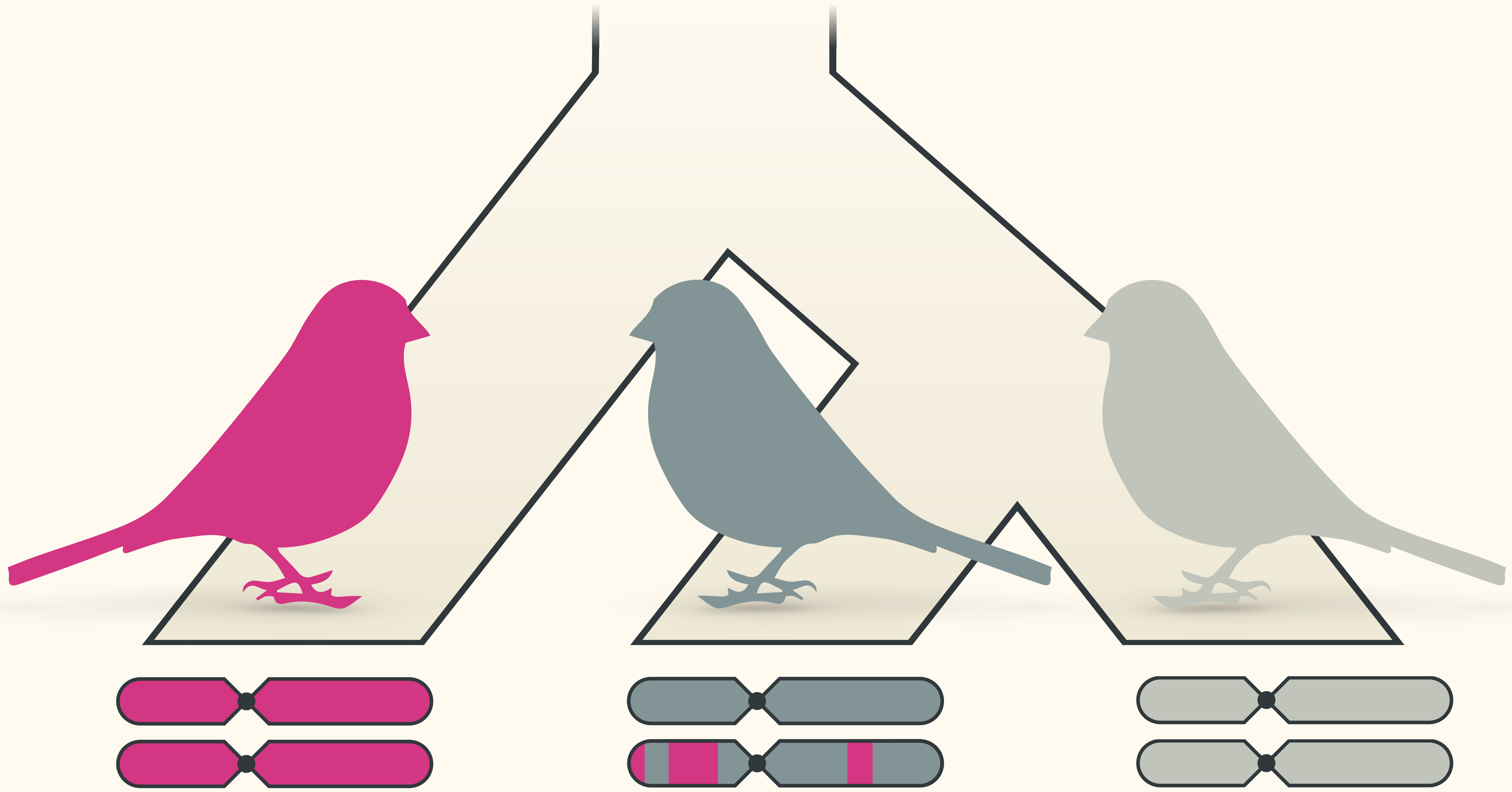
Introgression



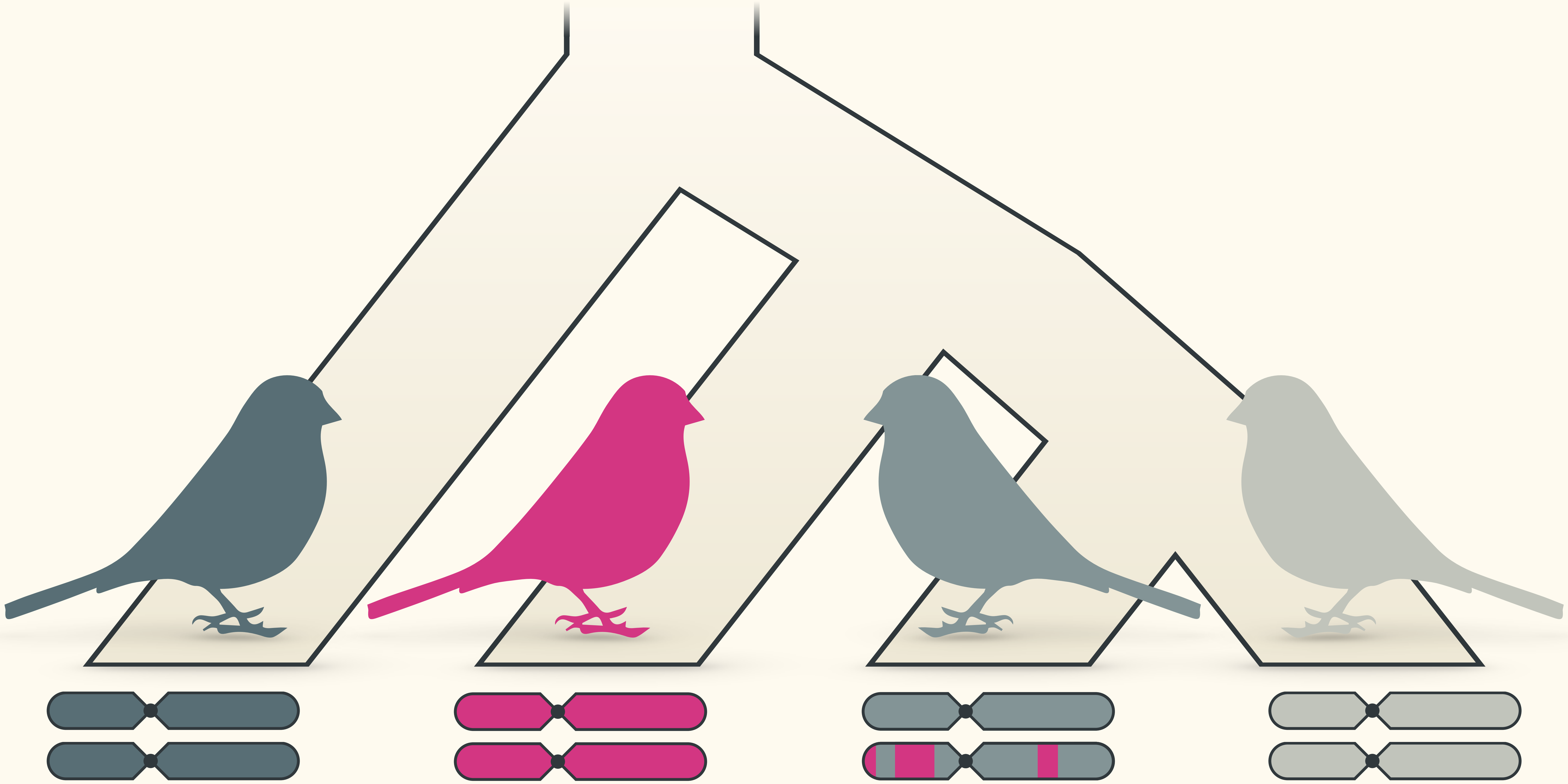
Introgression



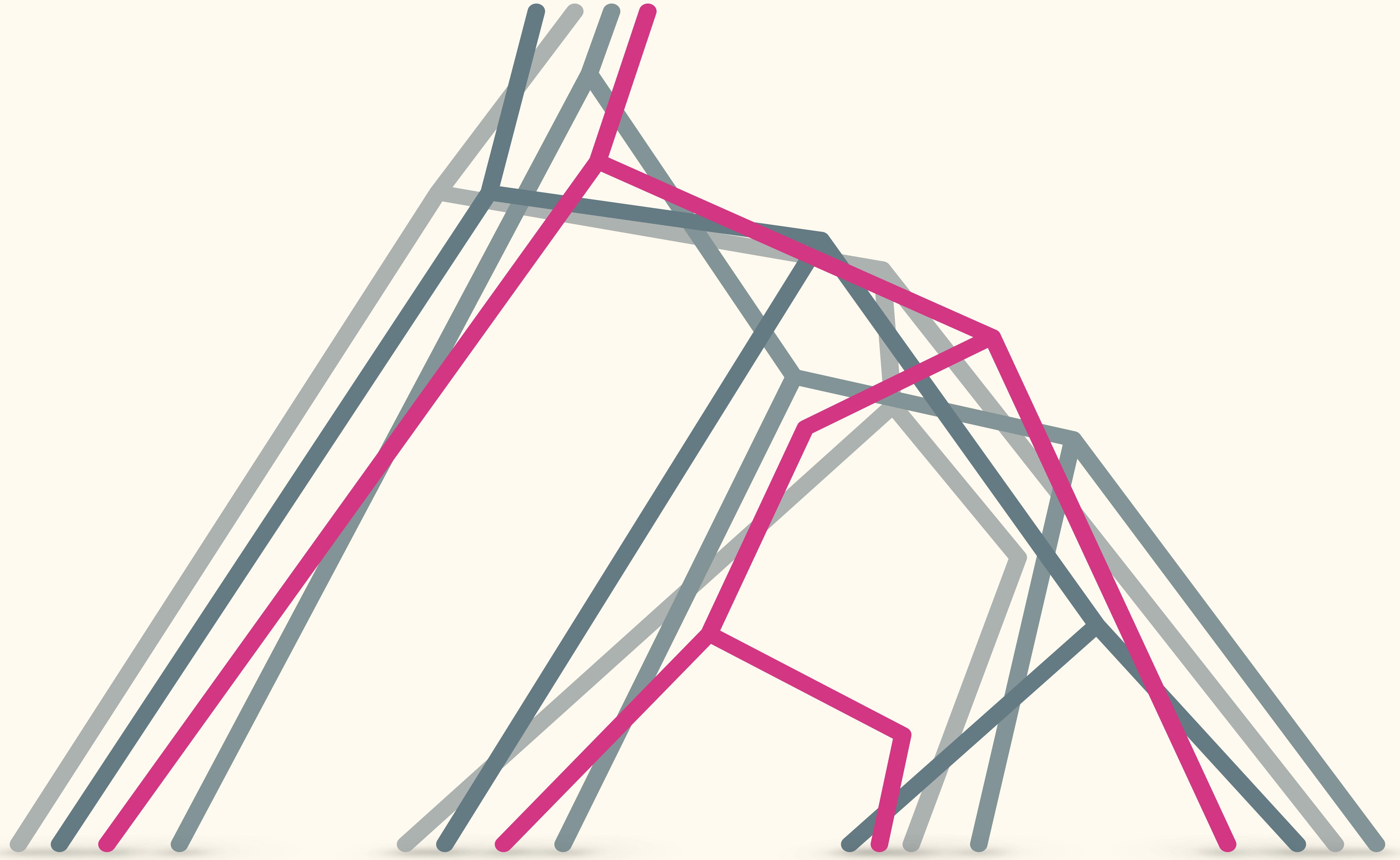
Introgression



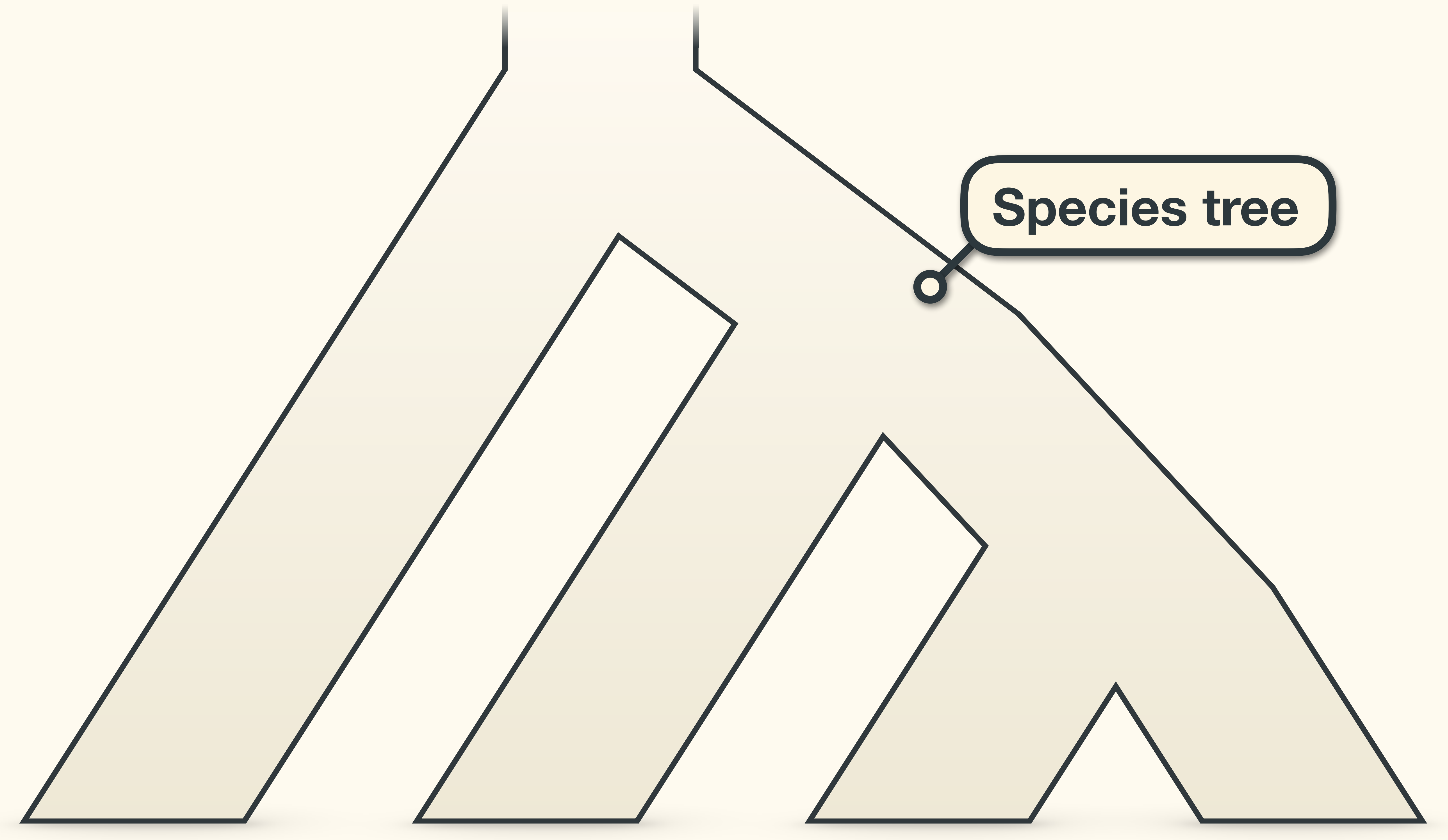
Detecting introgression



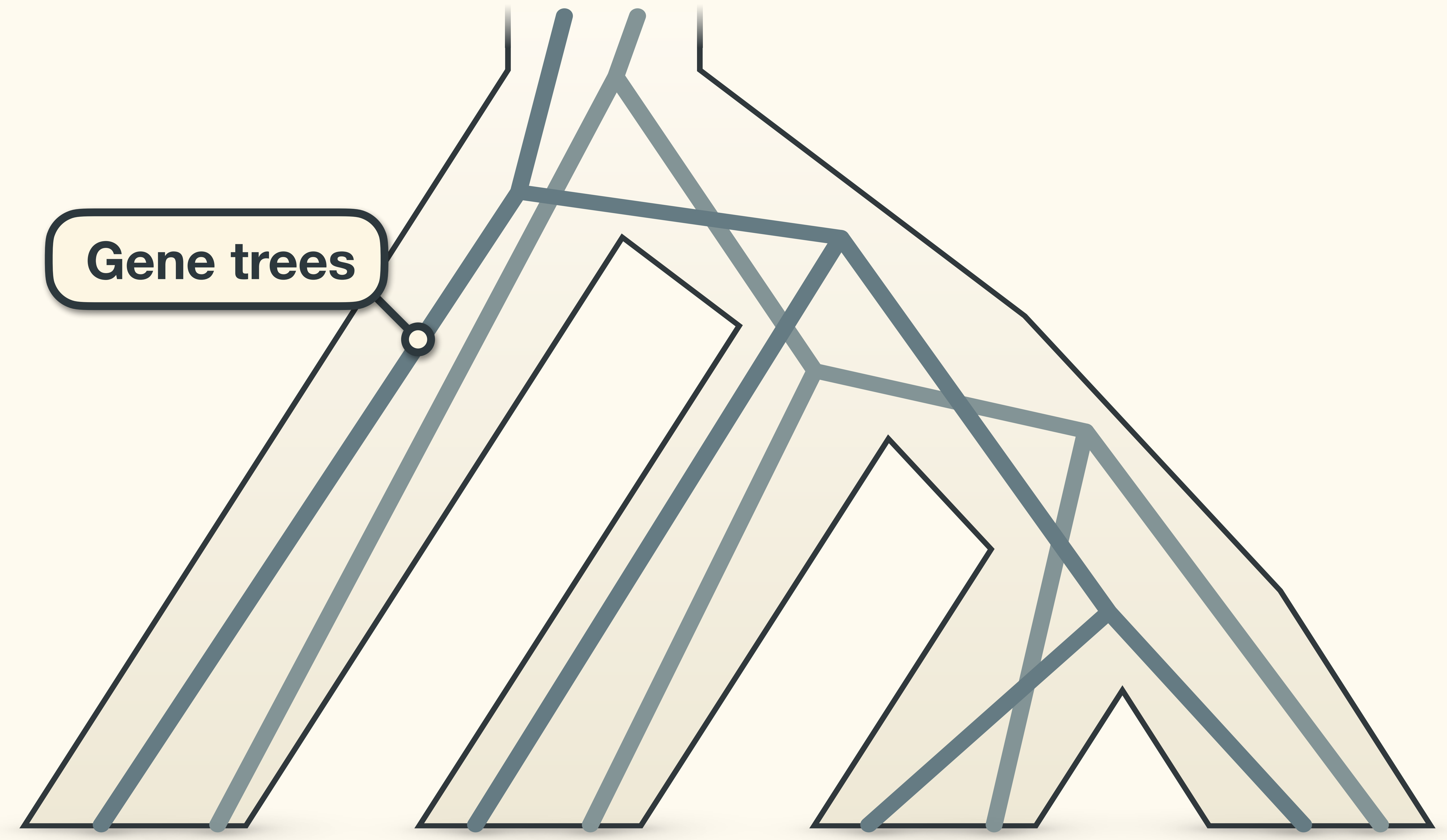
Gene trees



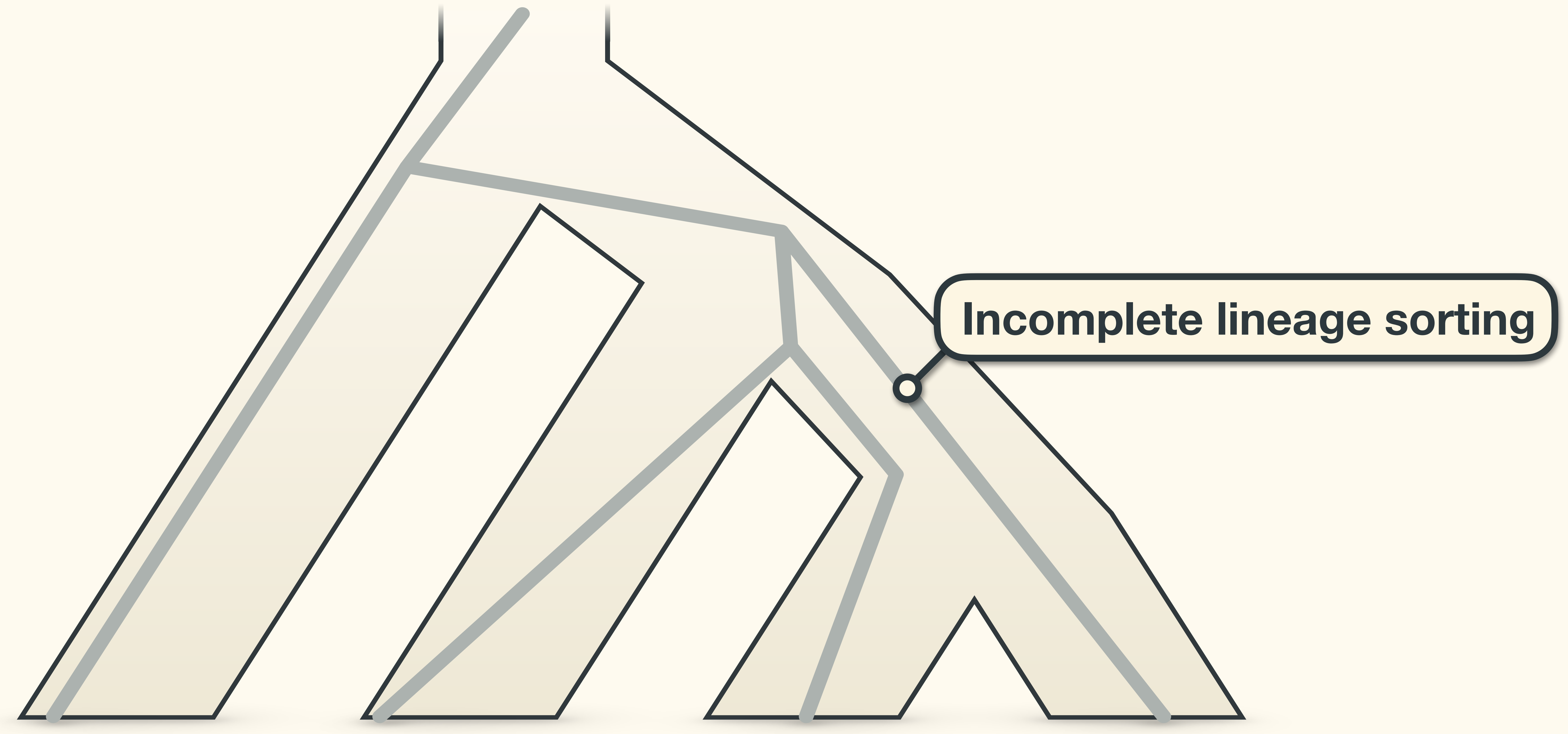
Gene trees



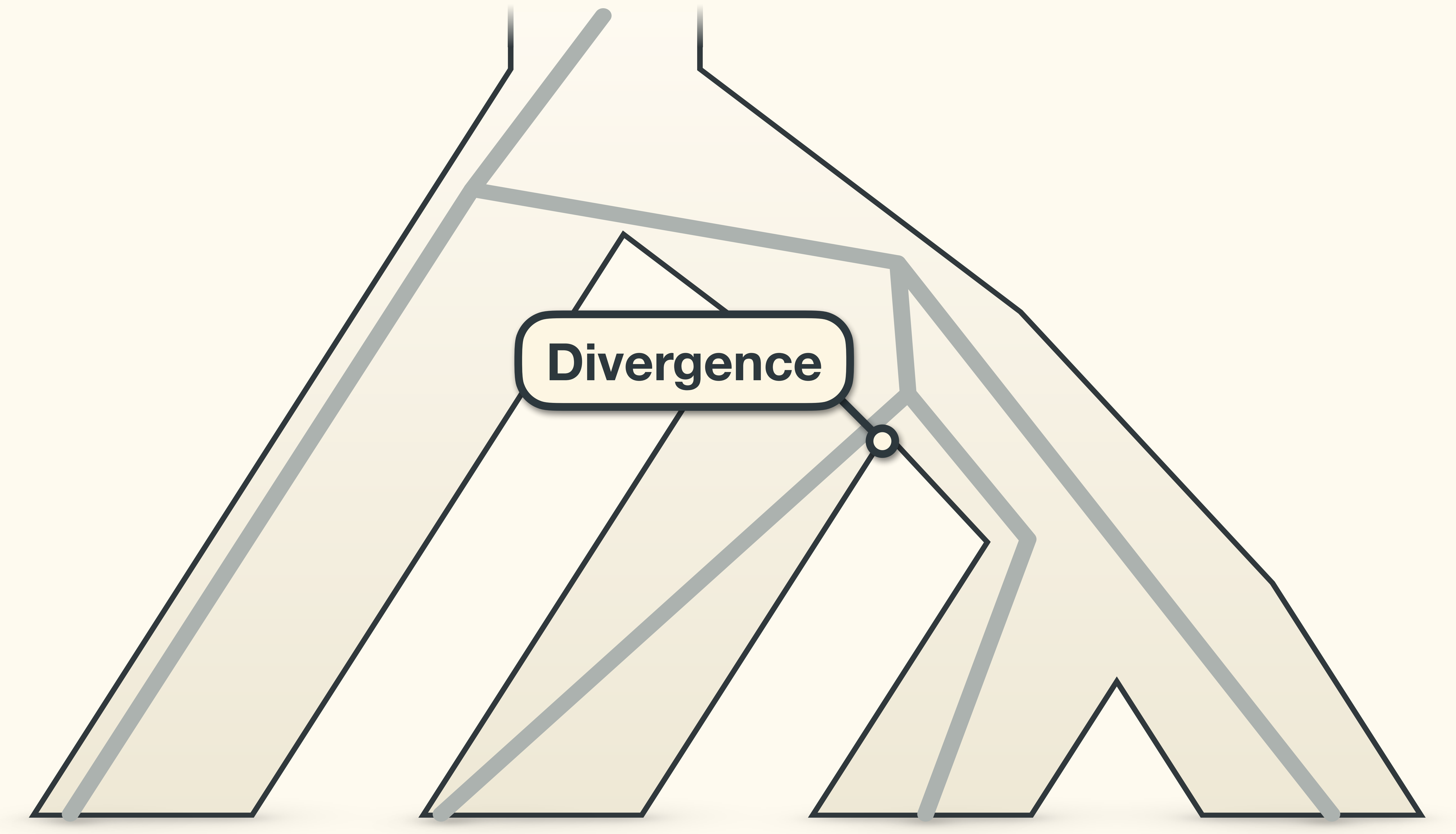
Gene trees



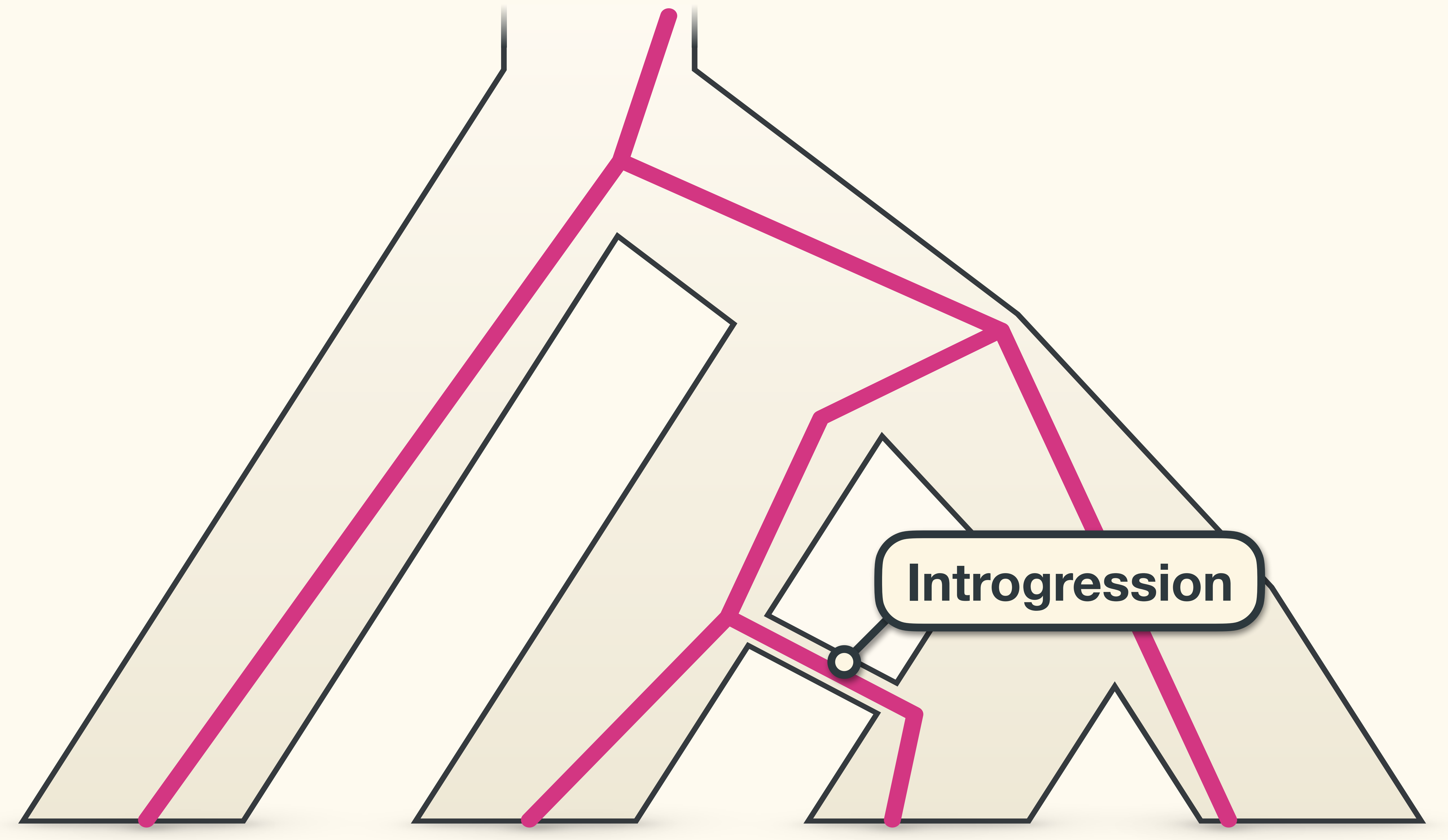
Gene trees



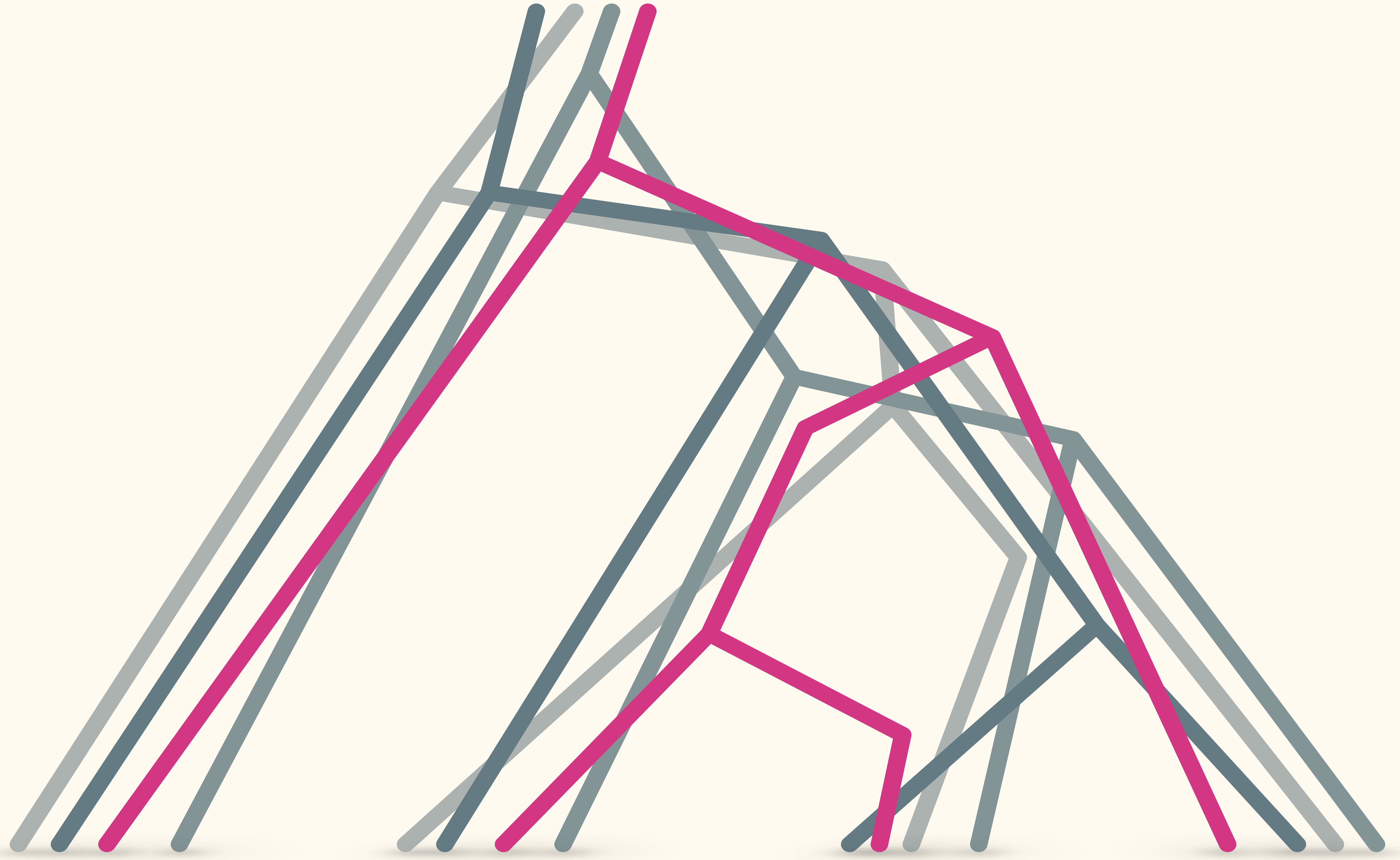
Gene trees



Gene trees

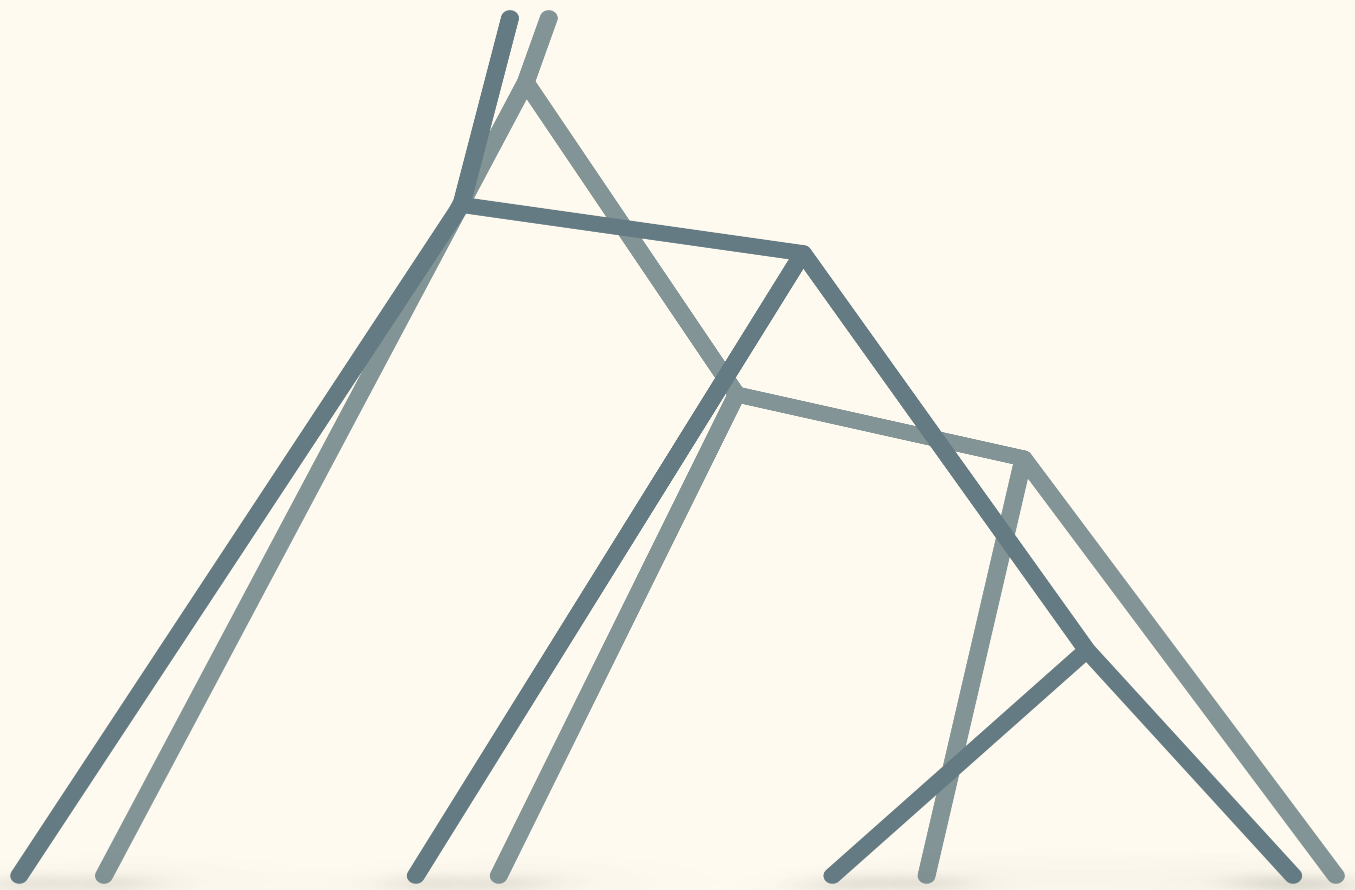


Gene trees

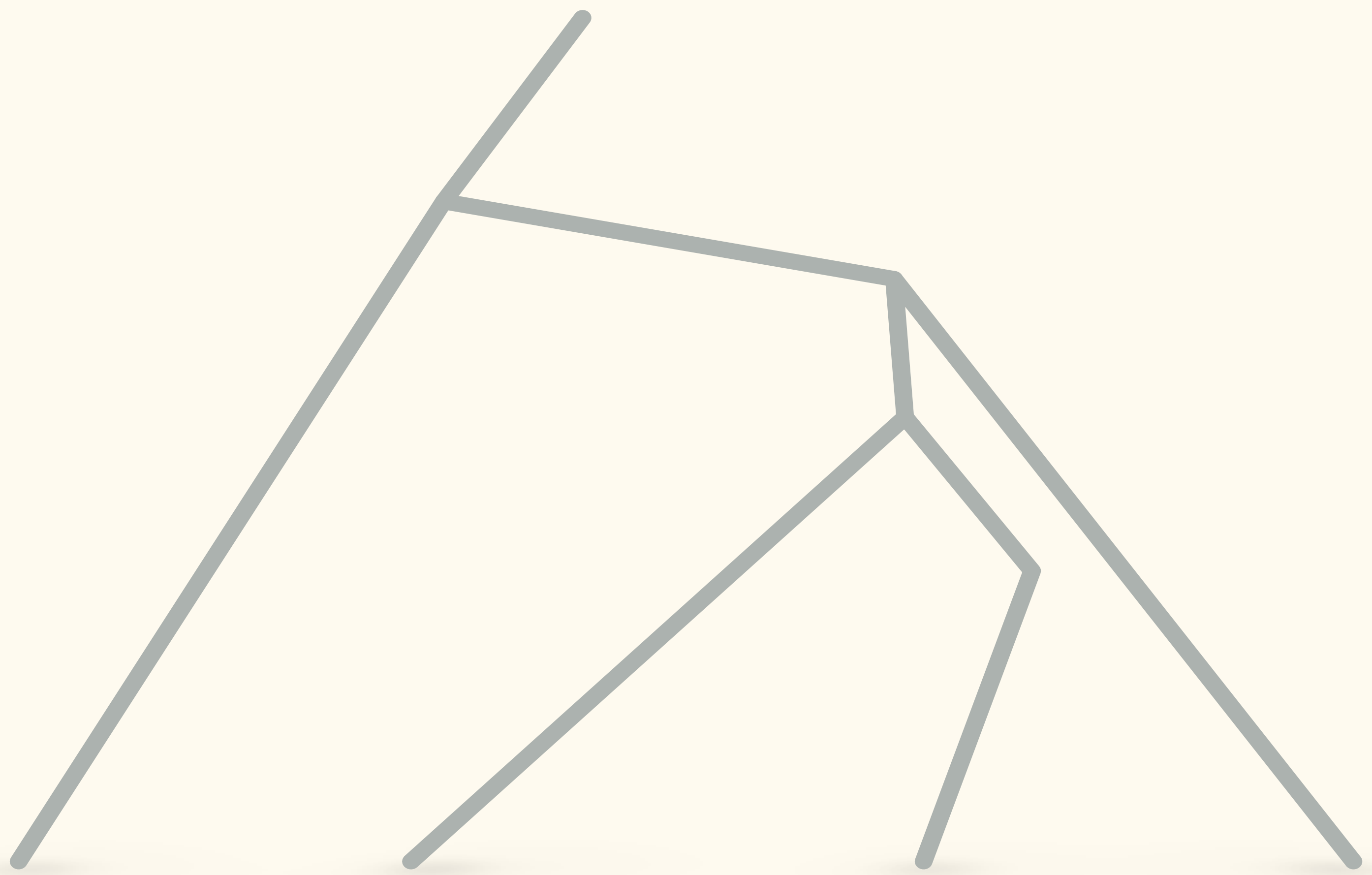


Gene trees

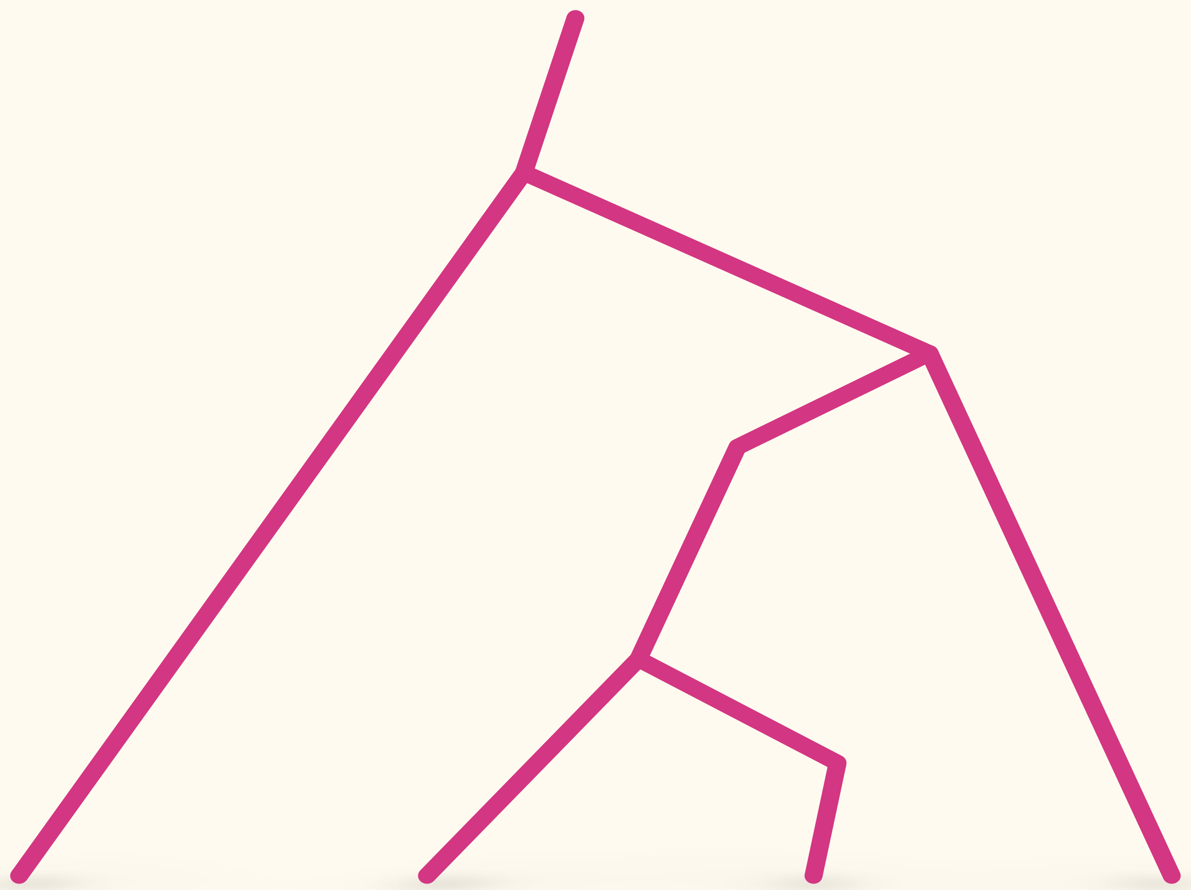
Species tree



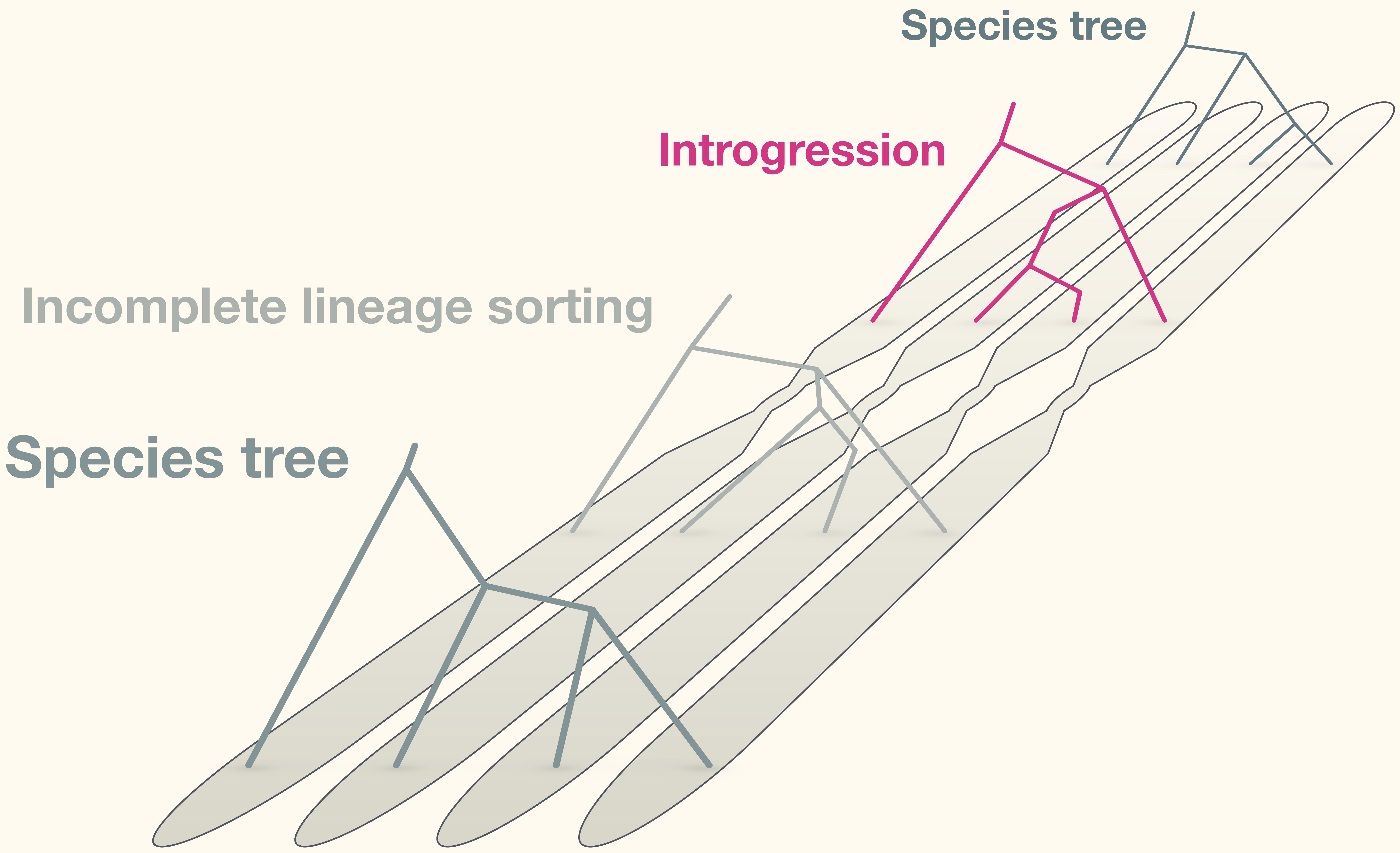
Incomplete lineage sorting



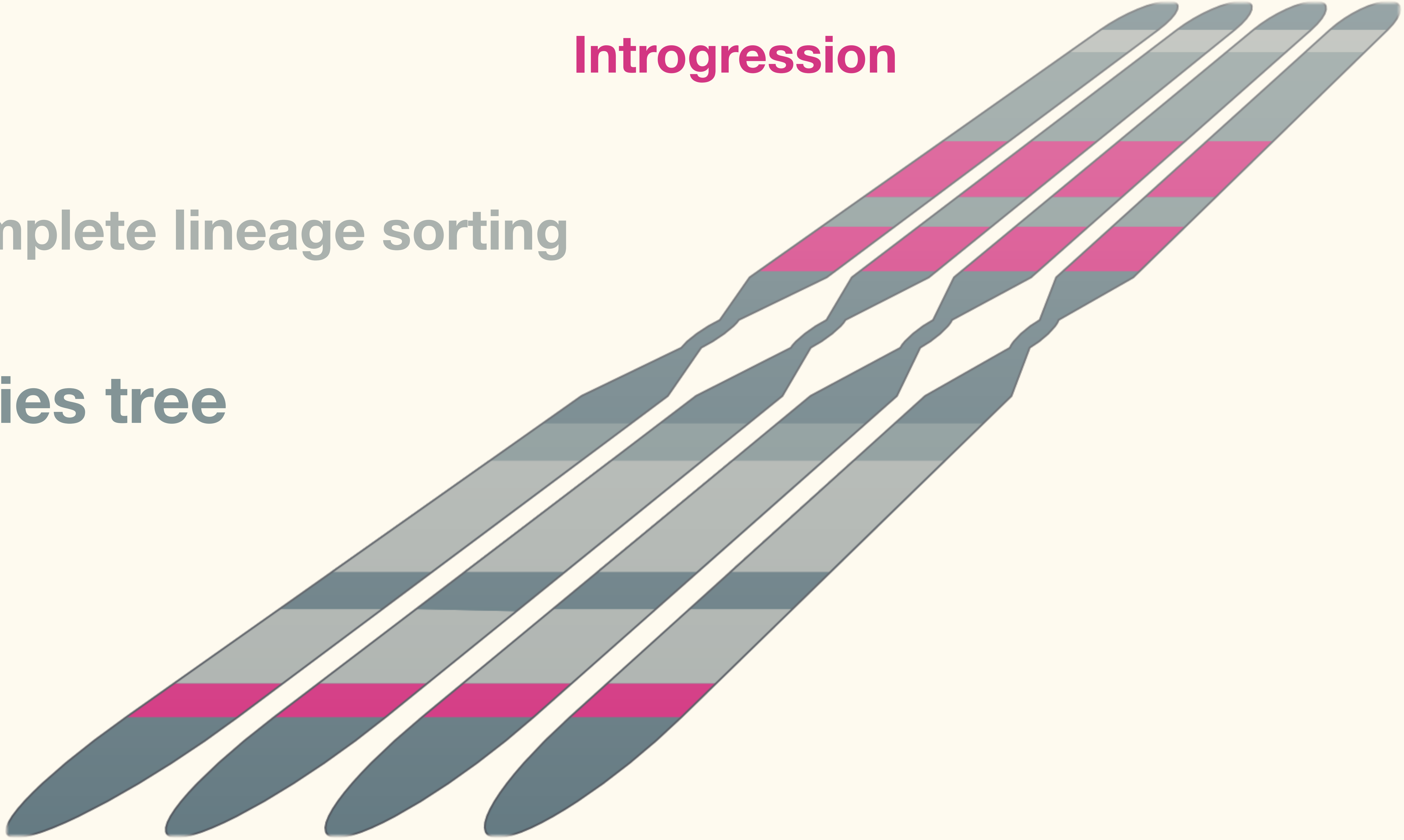
Introgression



Gene trees


Species tree
Introgression
Incomplete lineage sorting
Species tree

Gene trees



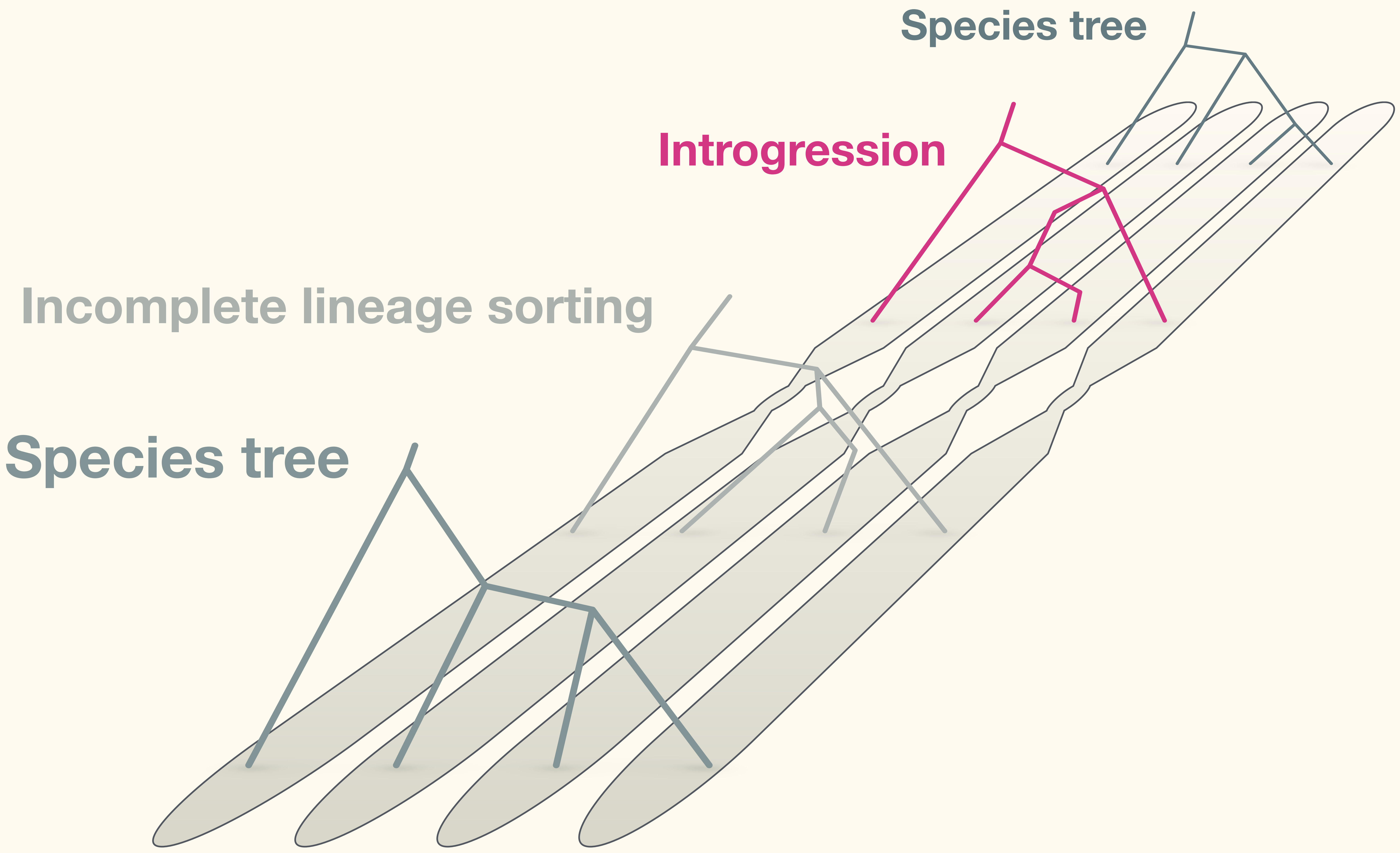
Species tree

Introgression

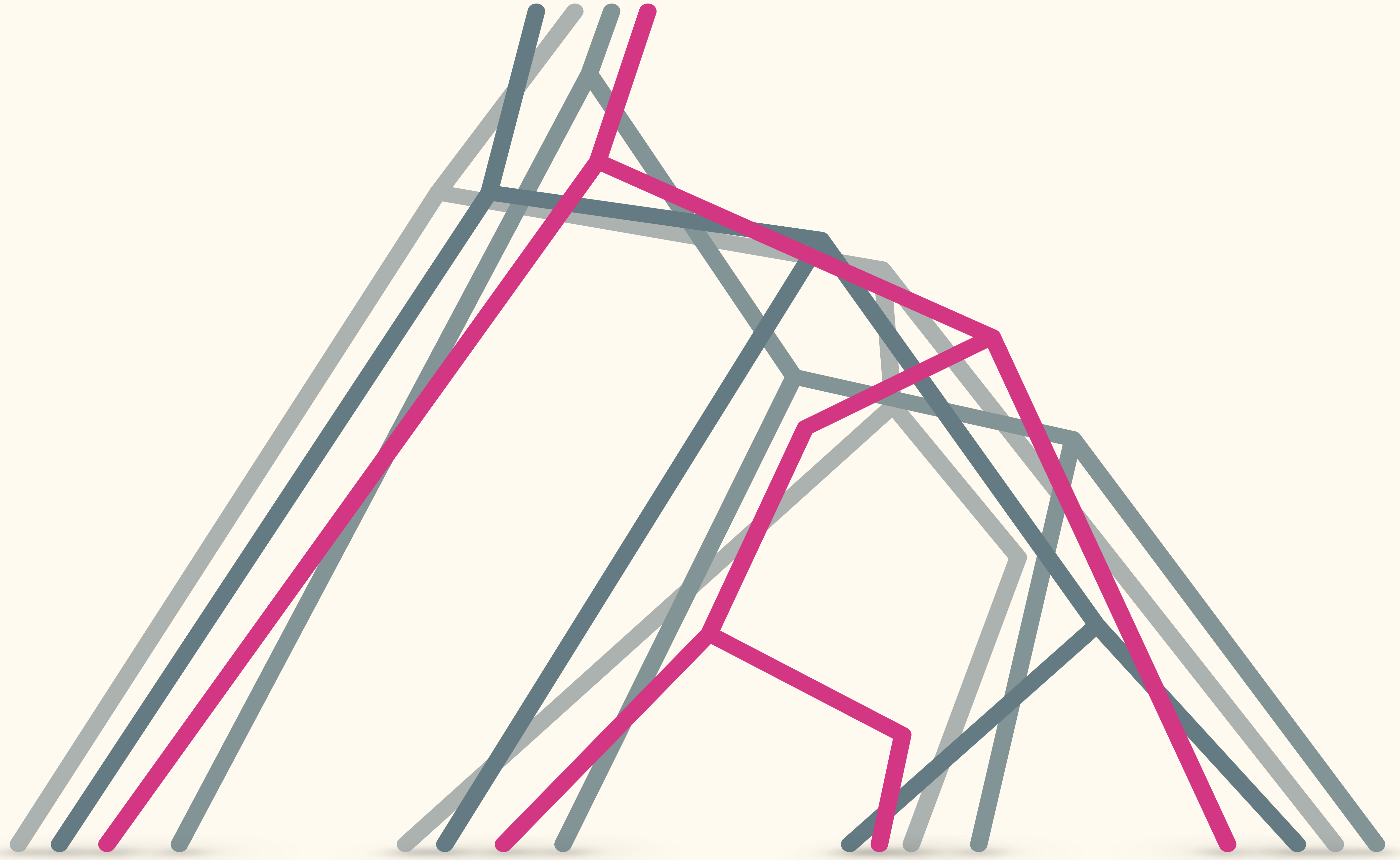
Incomplete lineage sorting

Species tree

Gene trees

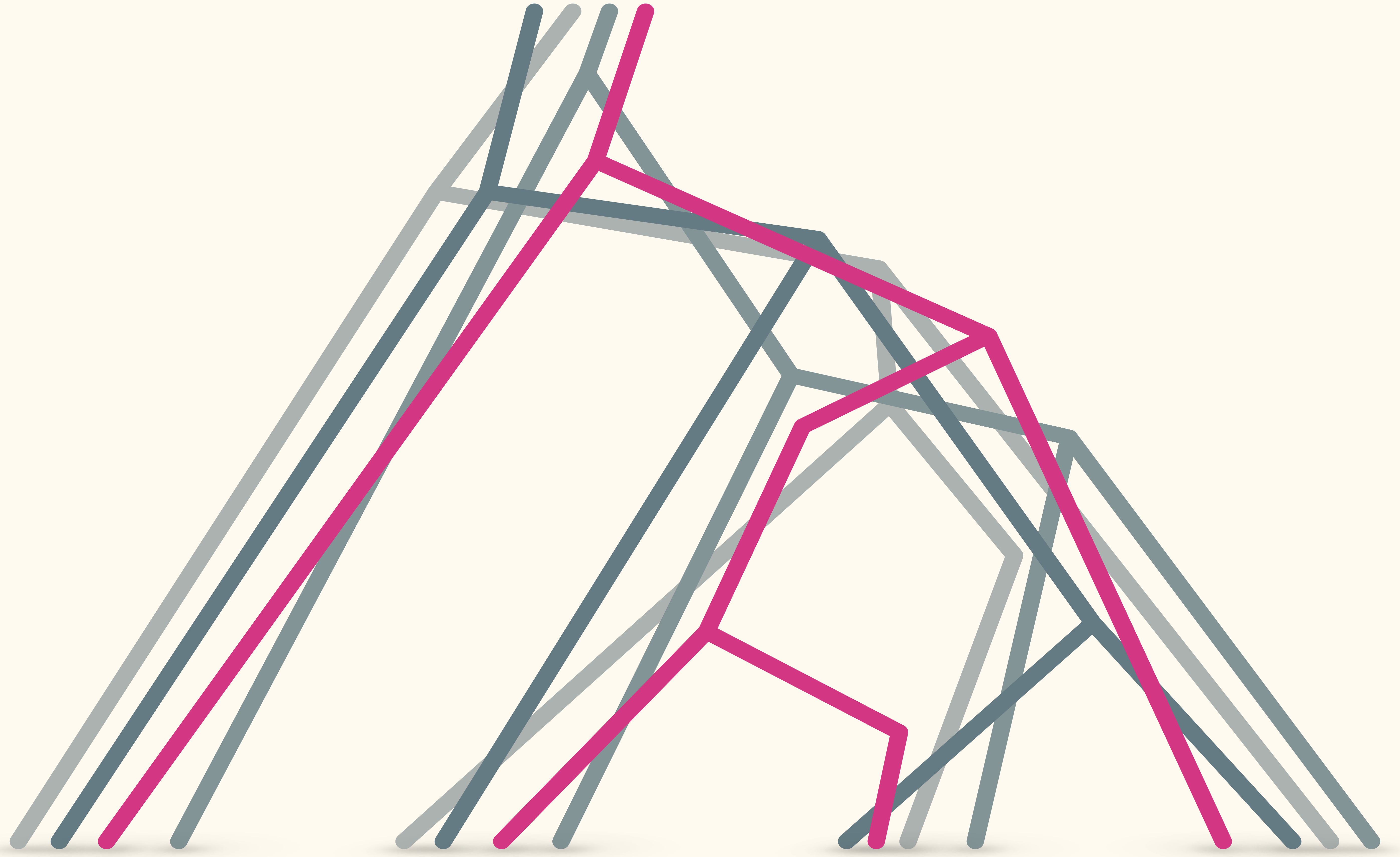

Species tree
Introgression
Incomplete lineage sorting
Species tree

Gene trees



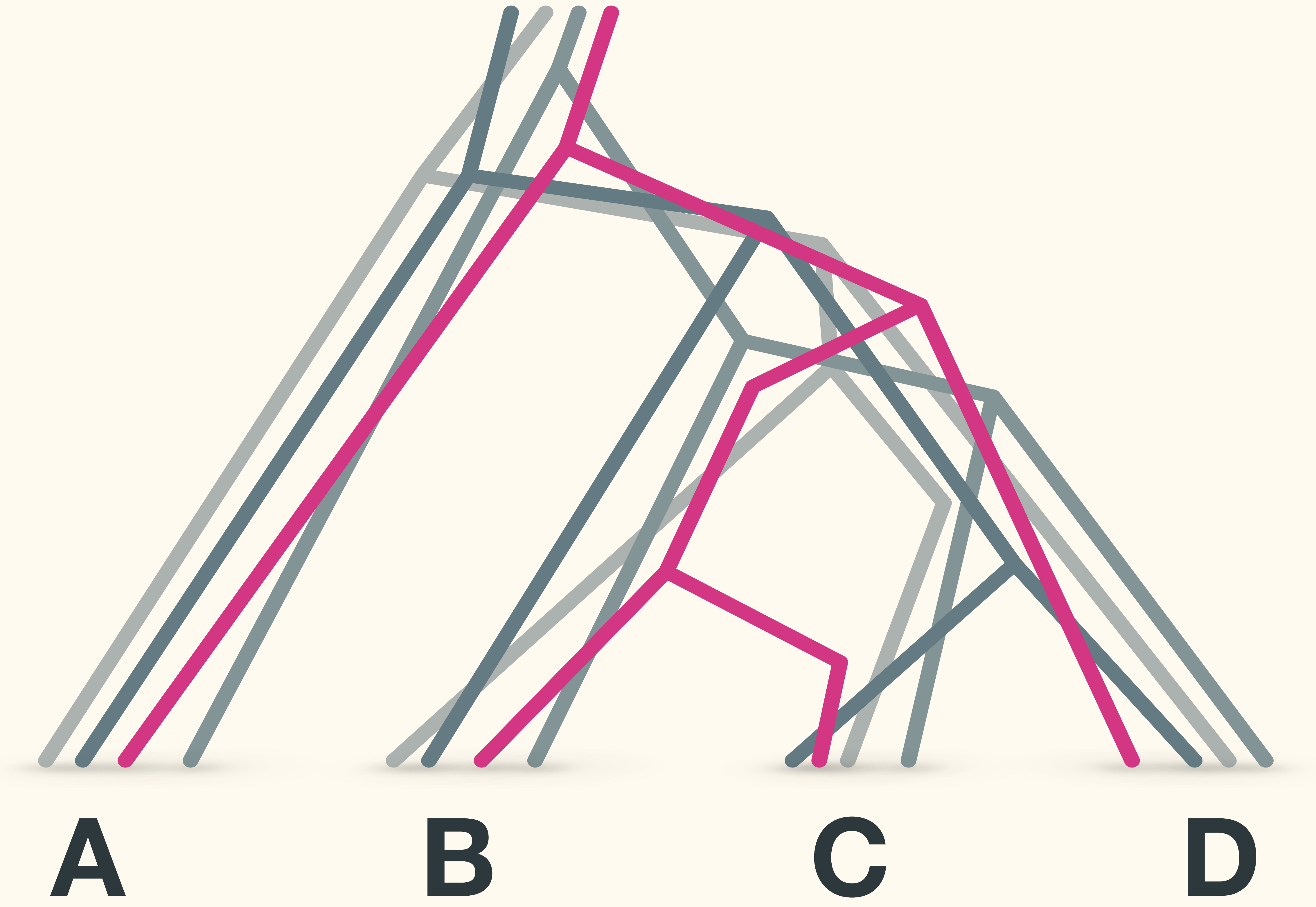
Introgression tests

Gene-tree asymmetry



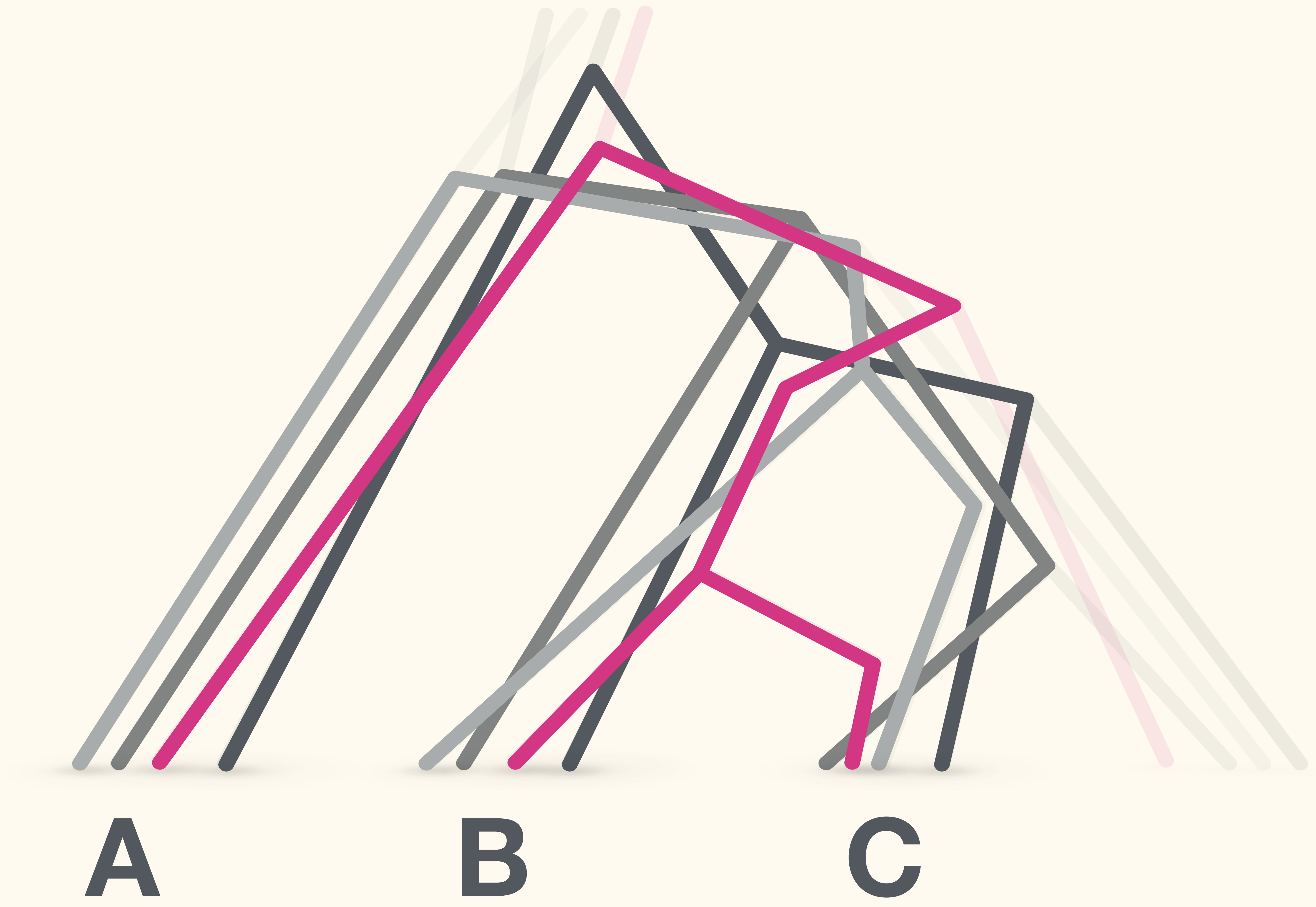
Introgression tests

Gene-tree asymmetry



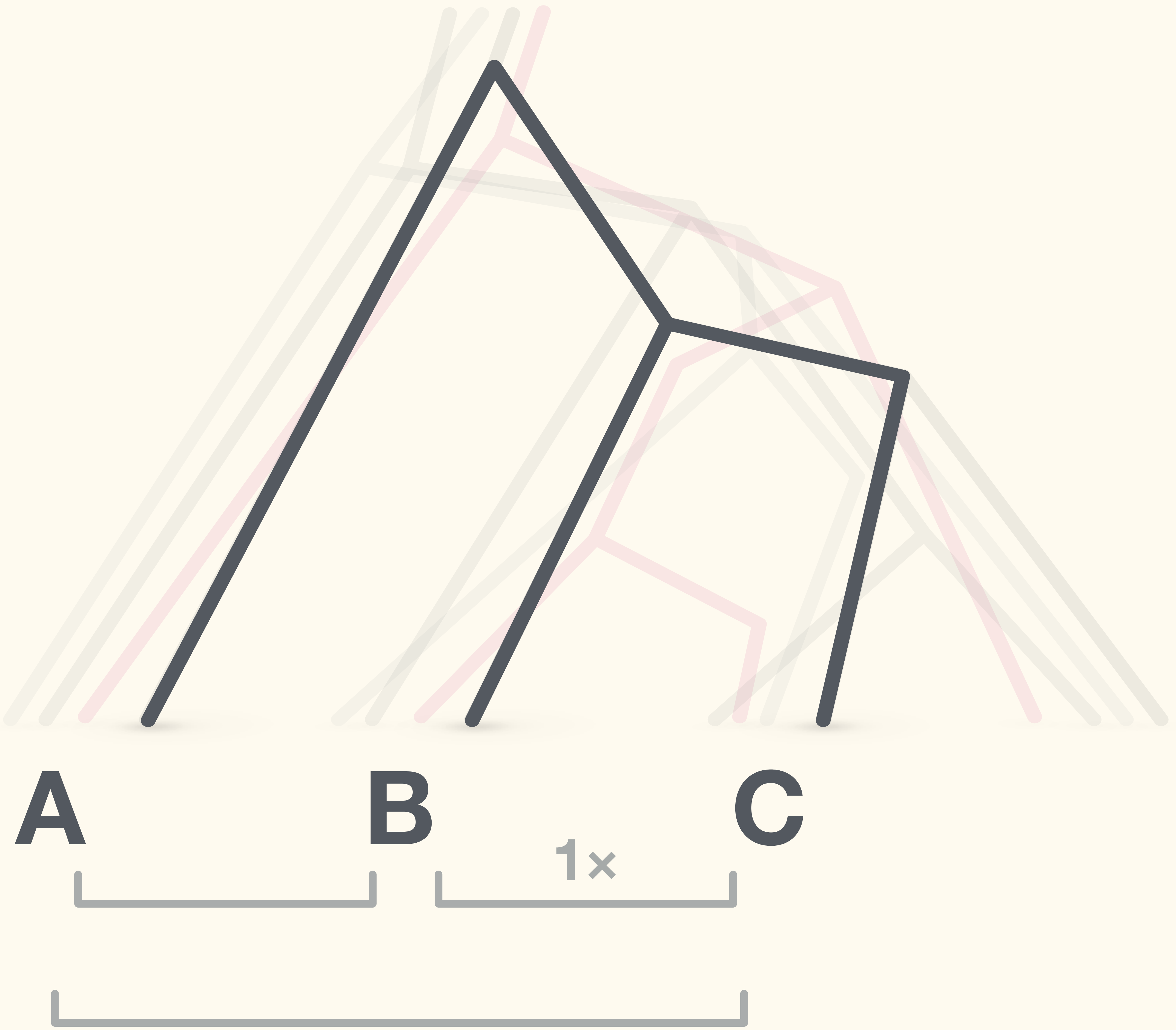
Introgression tests

Gene-tree asymmetry



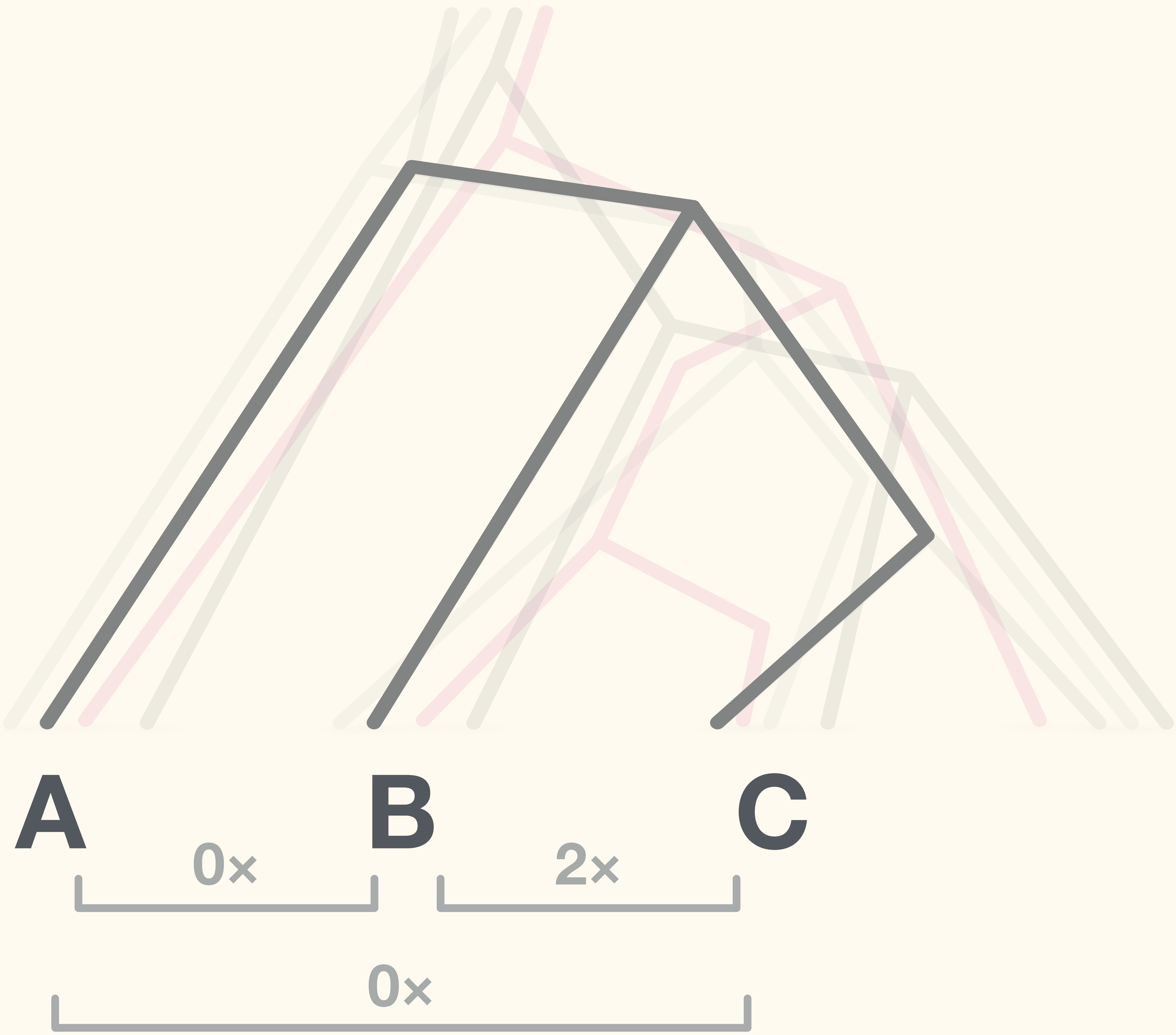
Introgression tests

Gene-tree asymmetry



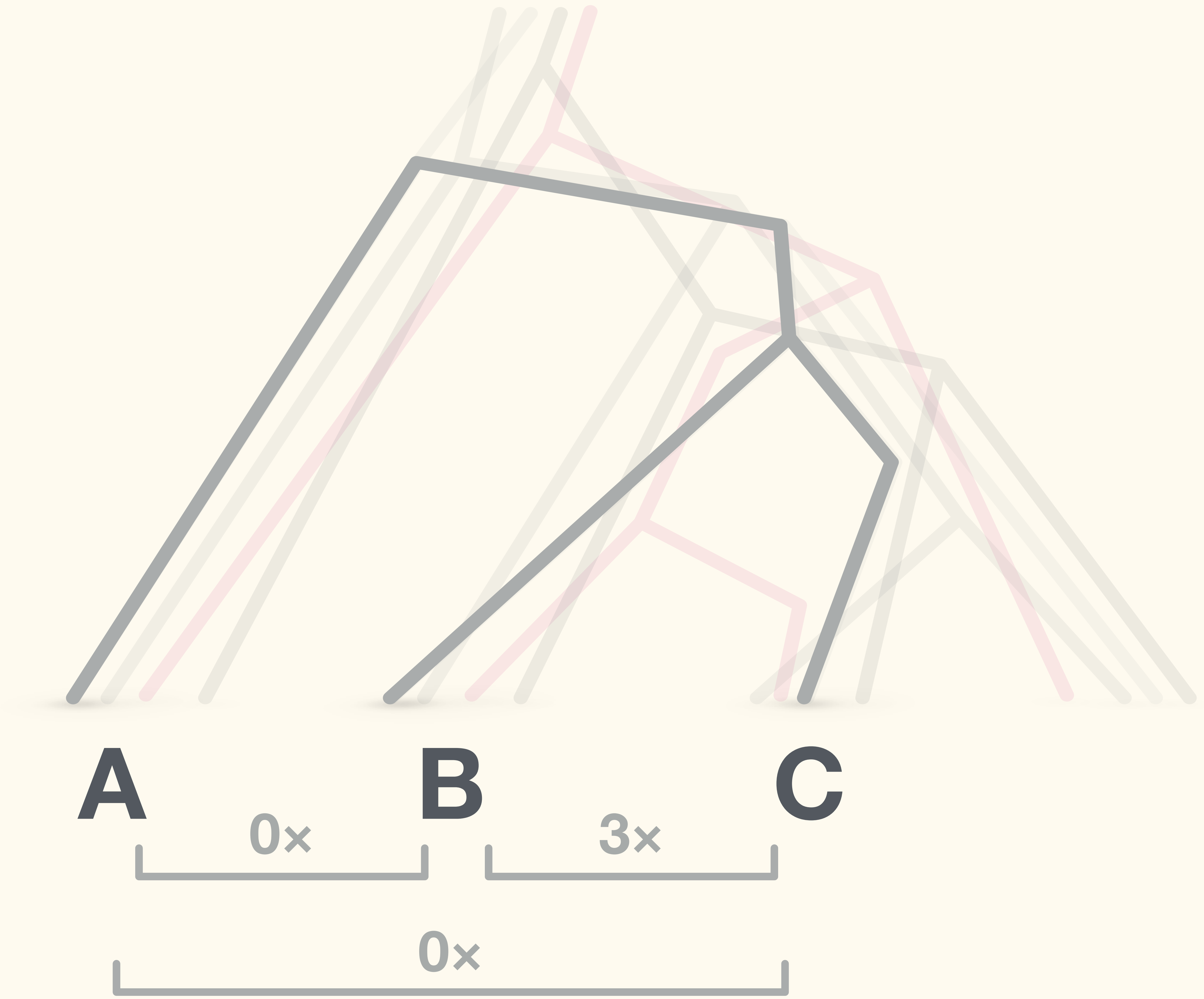
Introgression tests

Gene-tree asymmetry



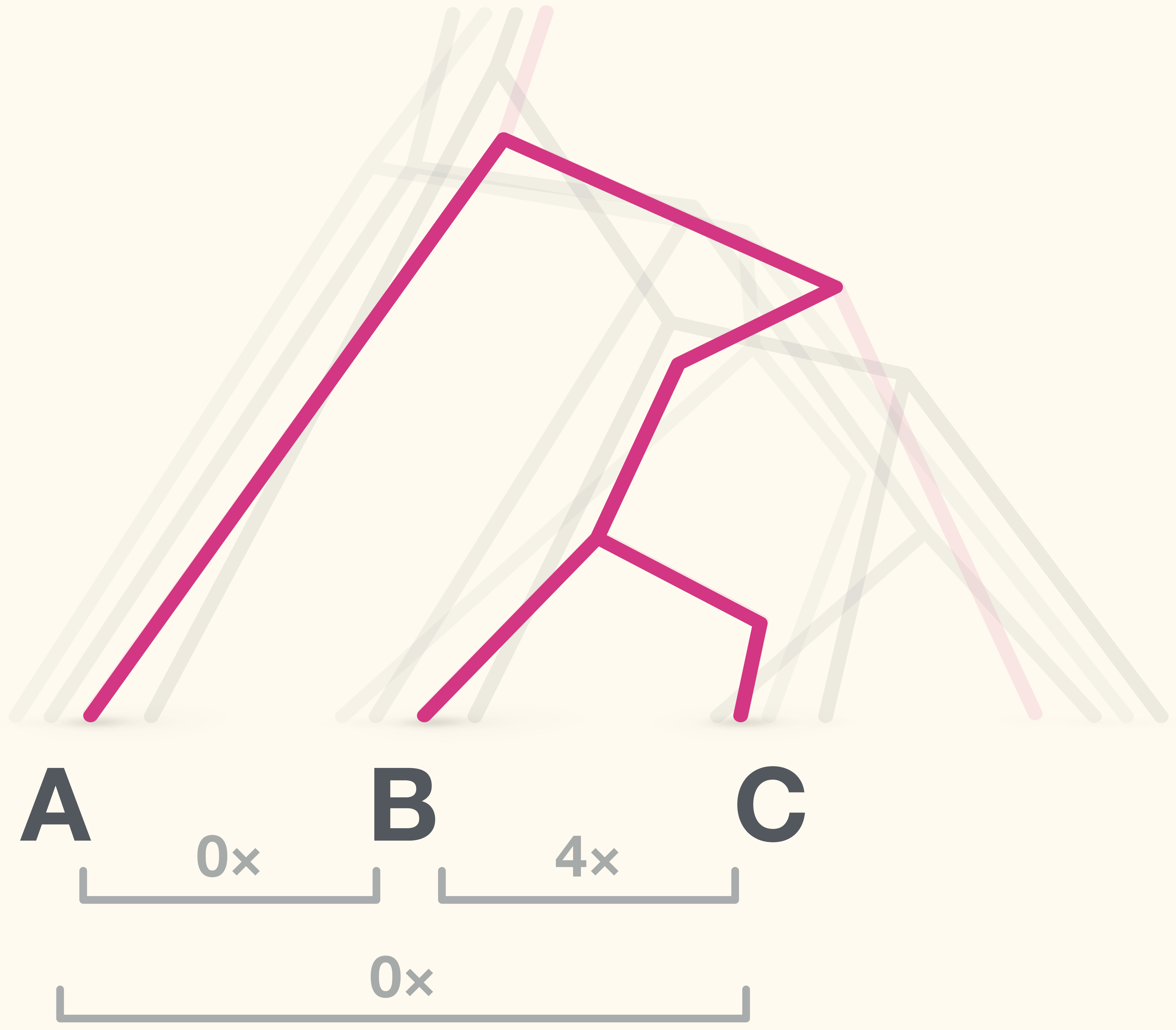
Introgression tests

Gene-tree asymmetry



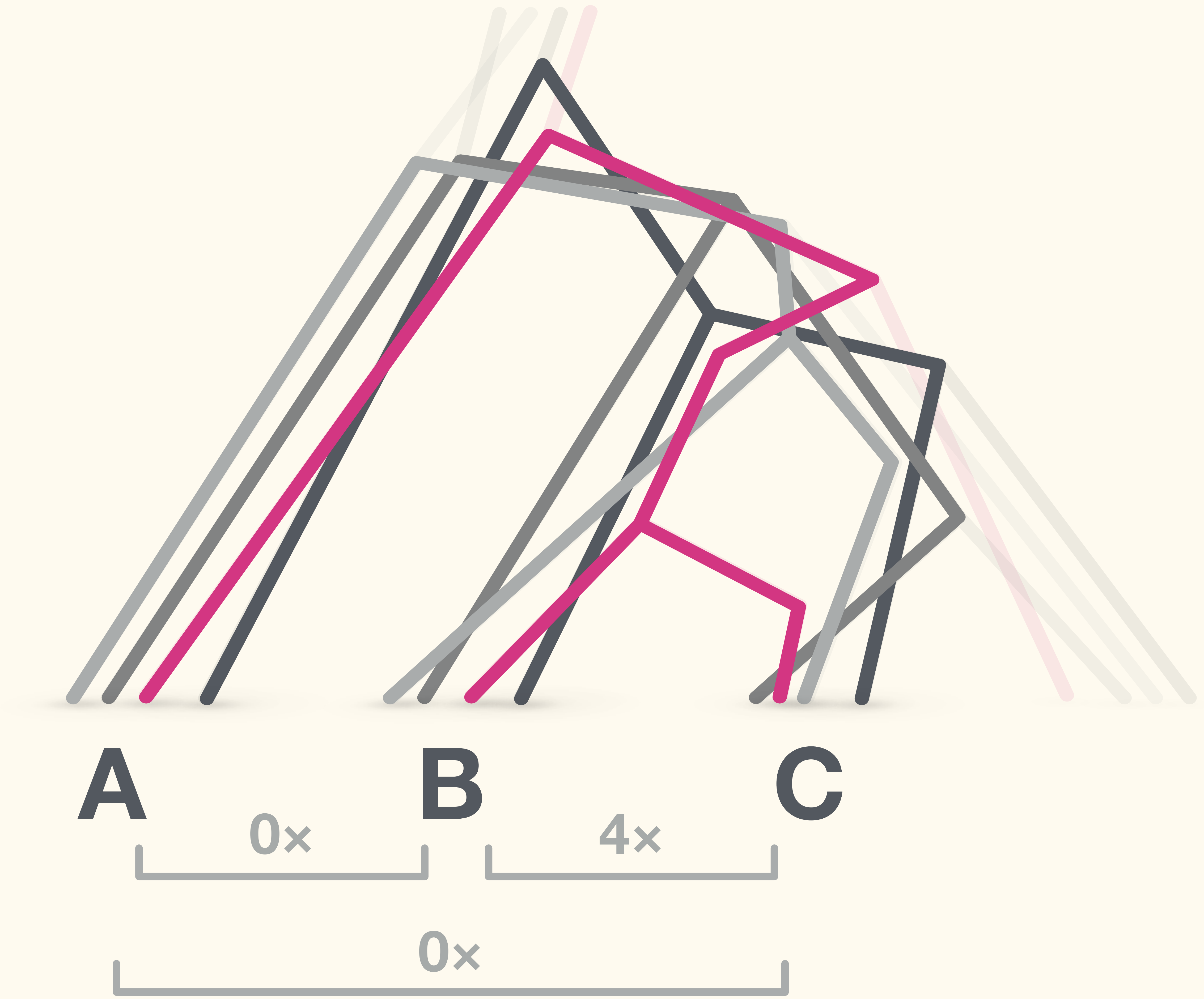
Introgression tests

Gene-tree asymmetry



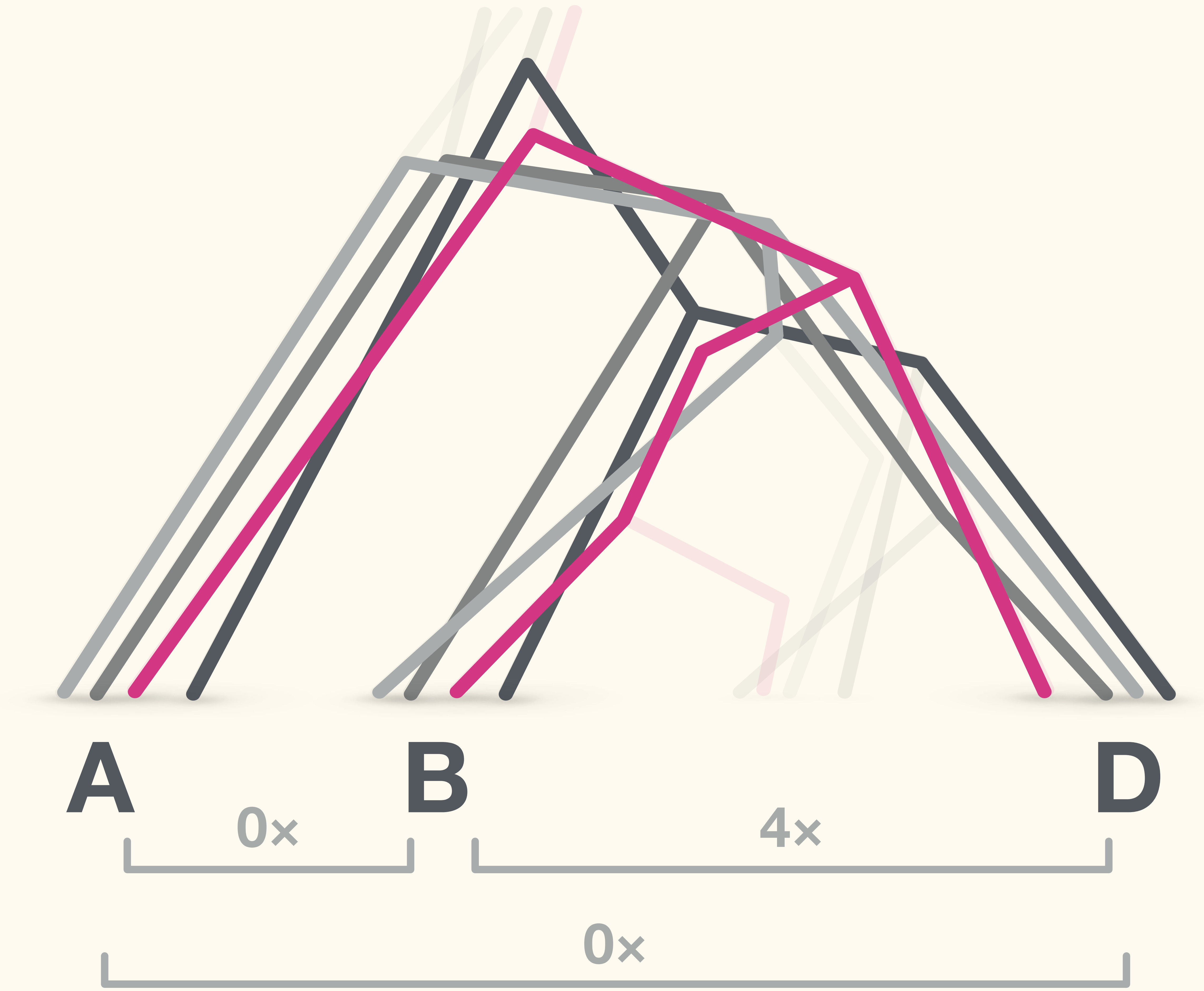
Introgression tests

Gene-tree asymmetry



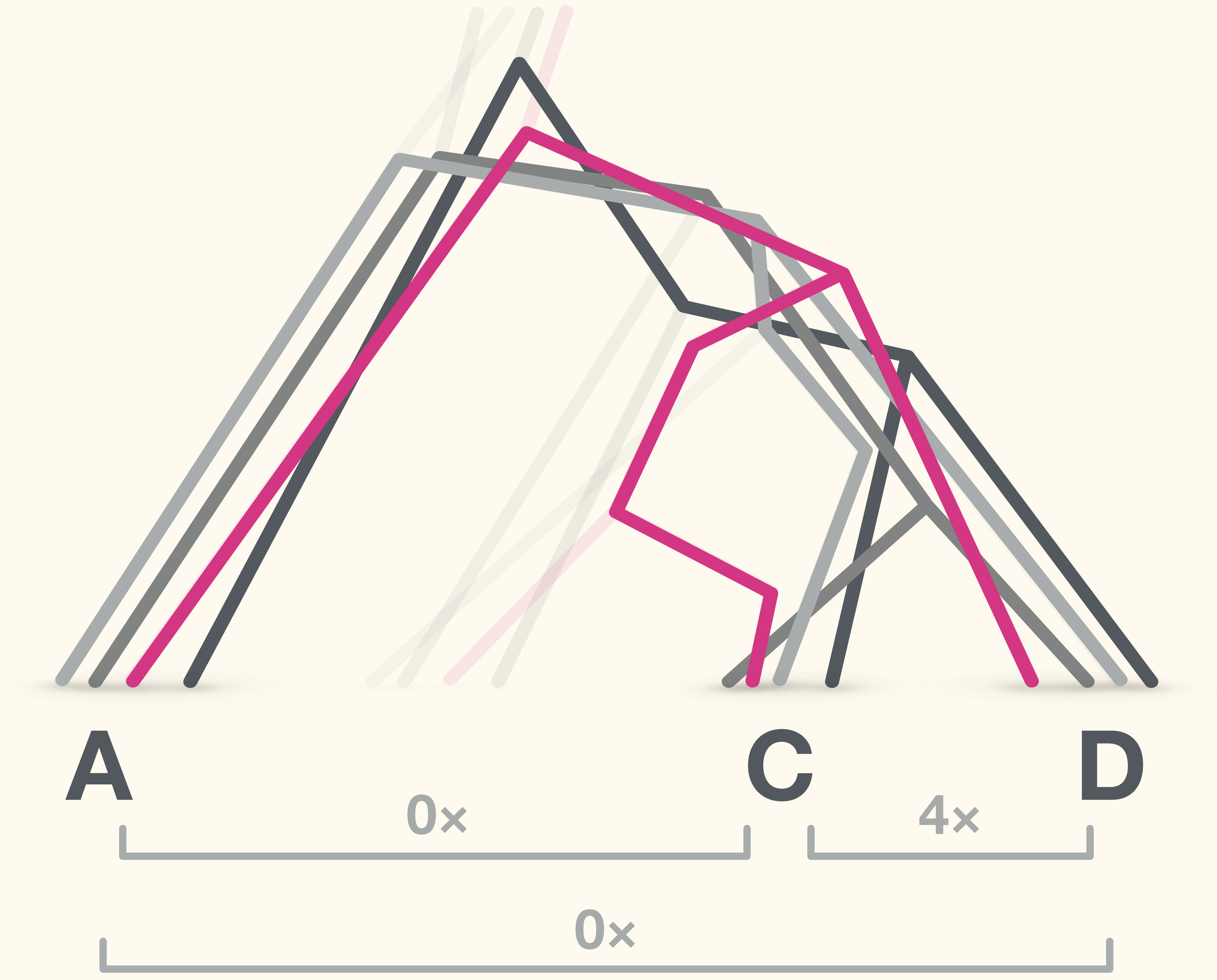
Introgression tests

Gene-tree asymmetry



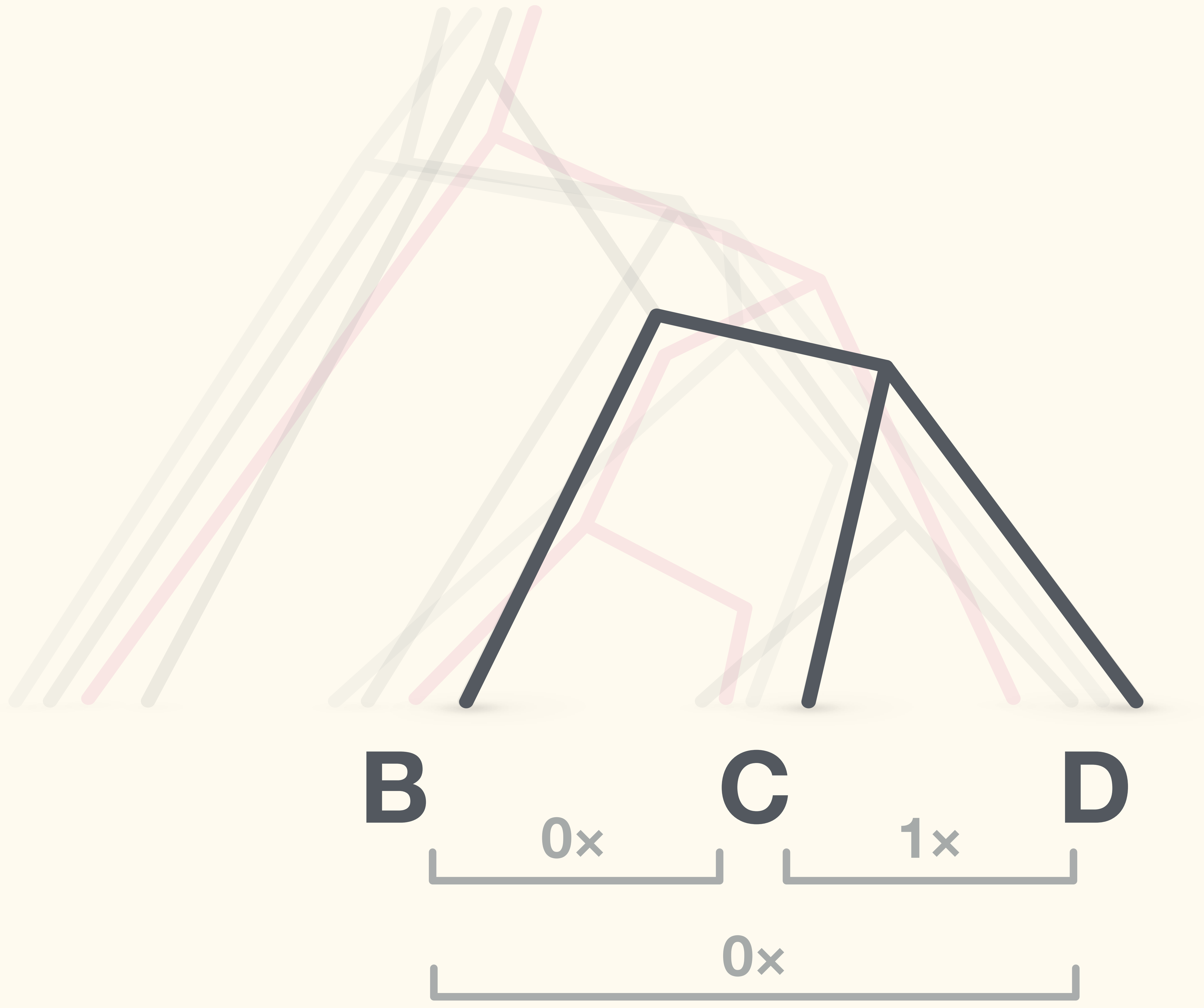
Introgression tests

Gene-tree asymmetry



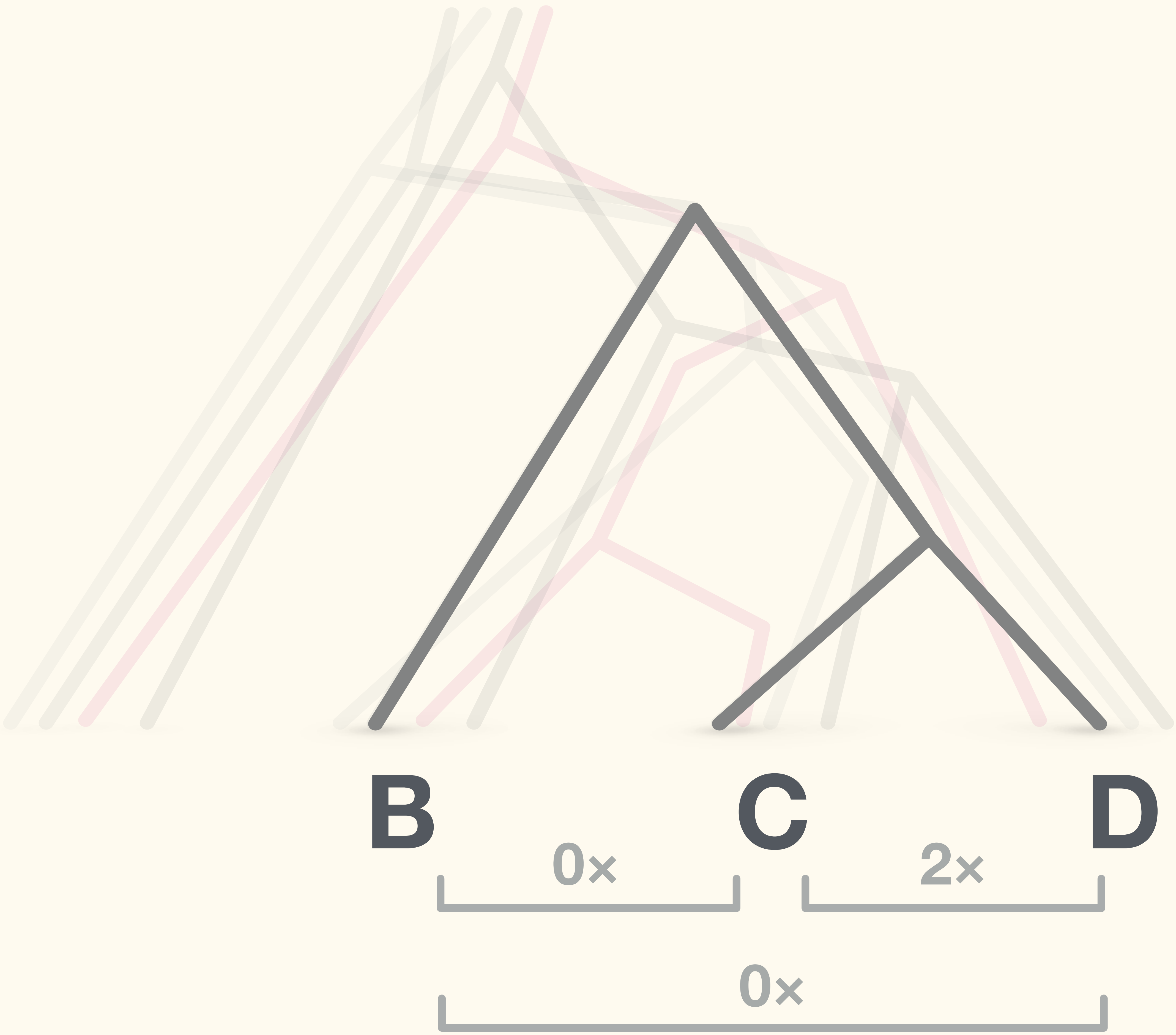
Introgression tests

Gene-tree asymmetry



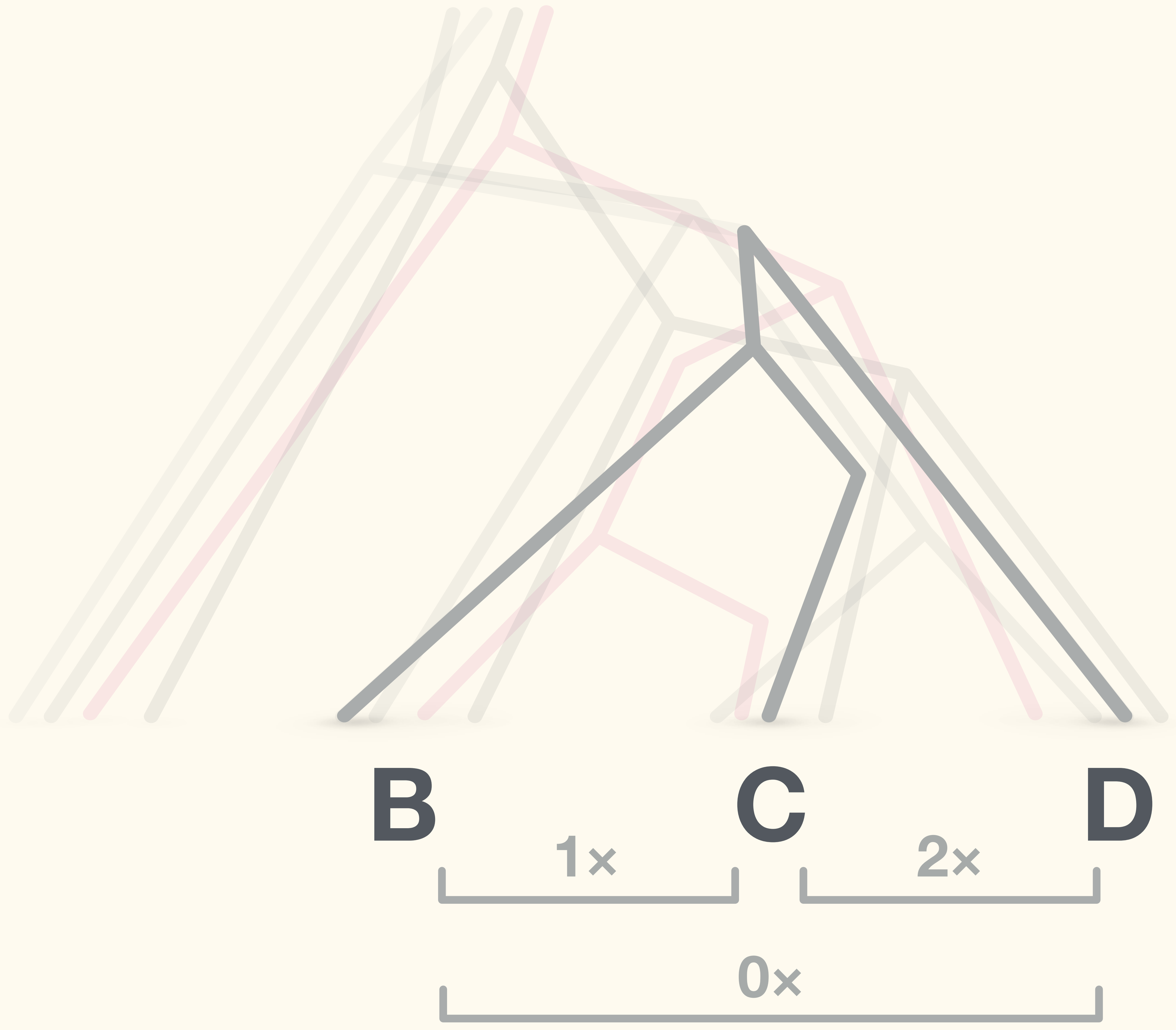
Introgression tests

Gene-tree asymmetry



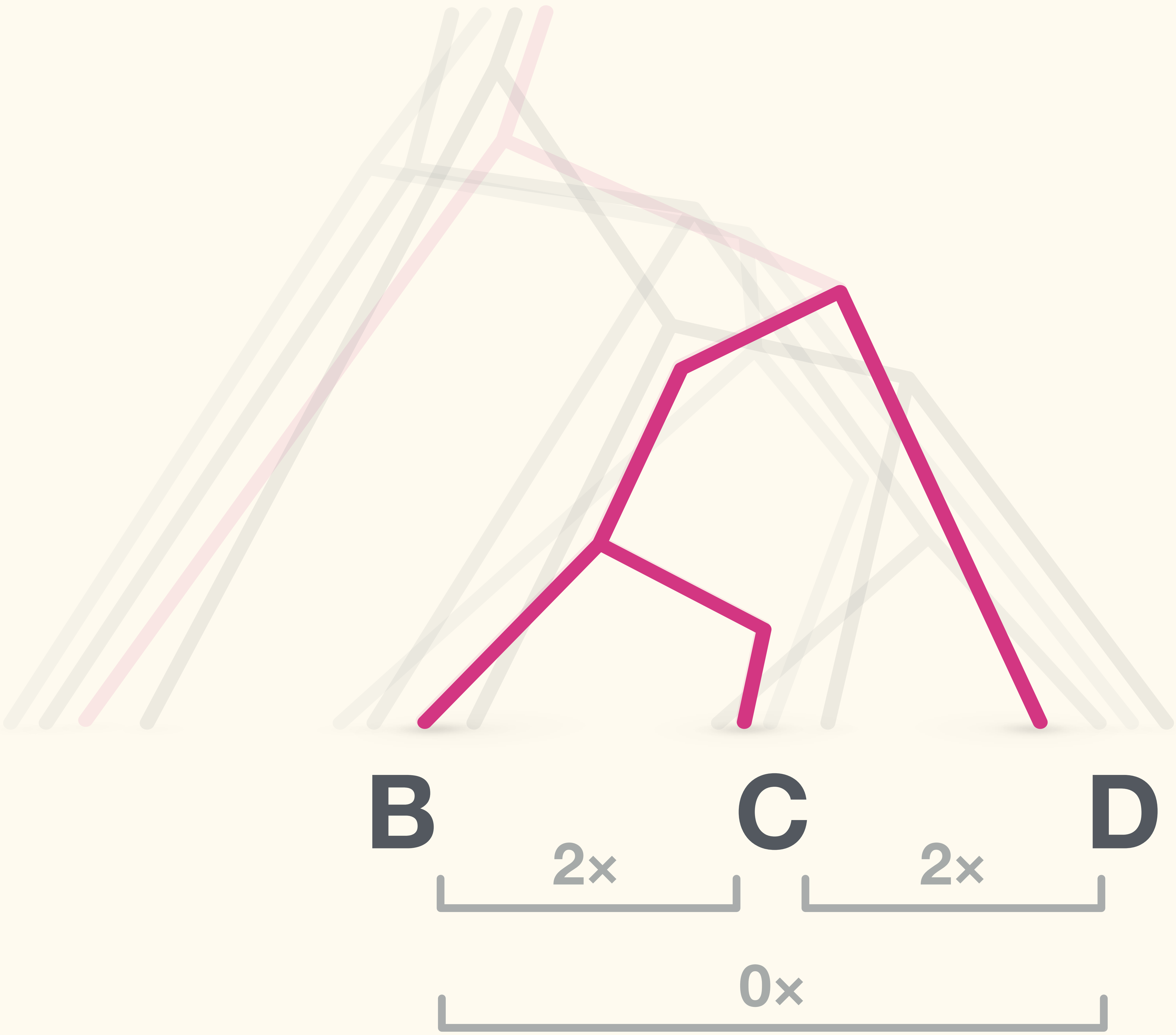
Introgression tests

Gene-tree asymmetry



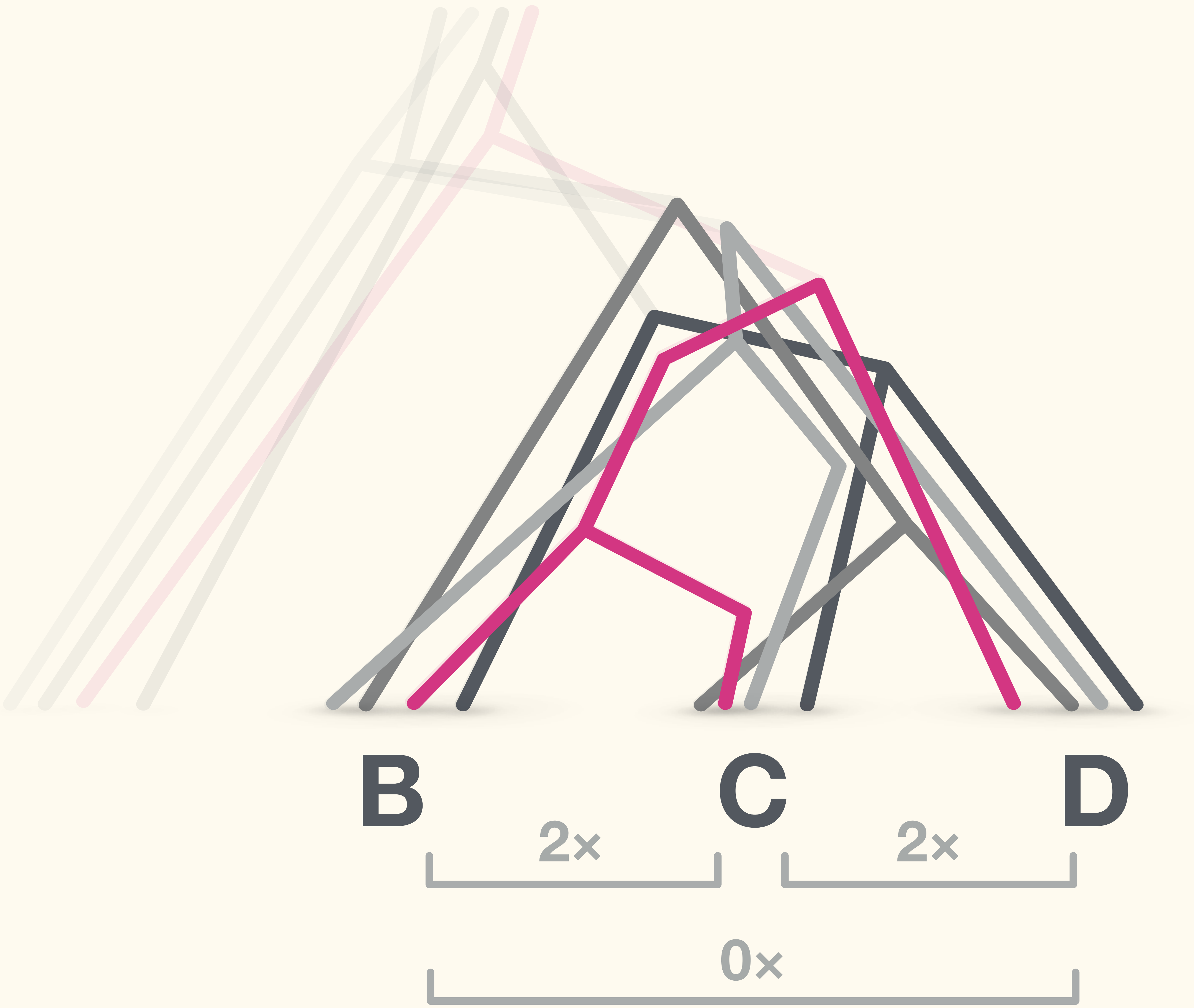
Introgression tests

Gene-tree asymmetry



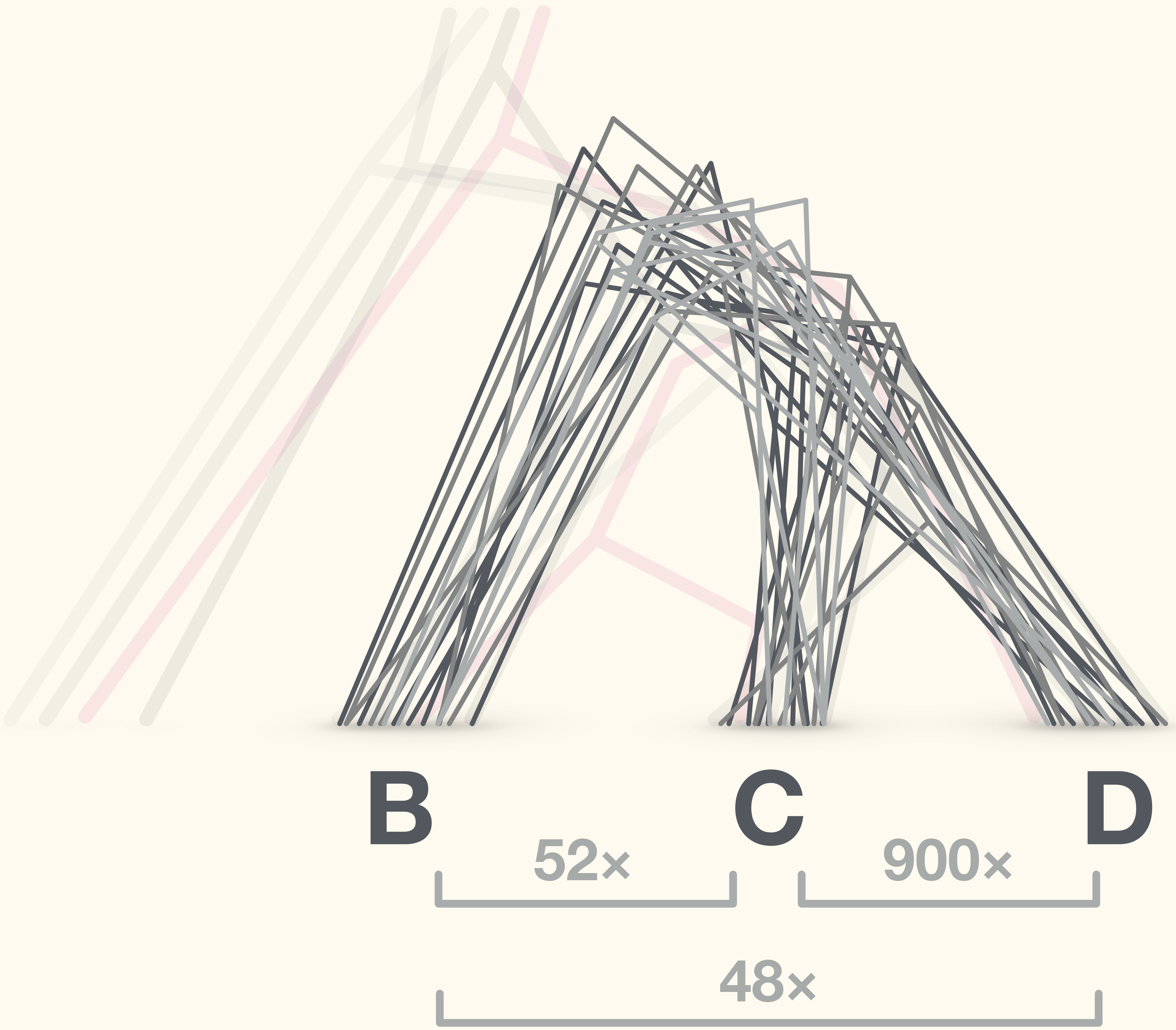
Introgression tests

Gene-tree asymmetry



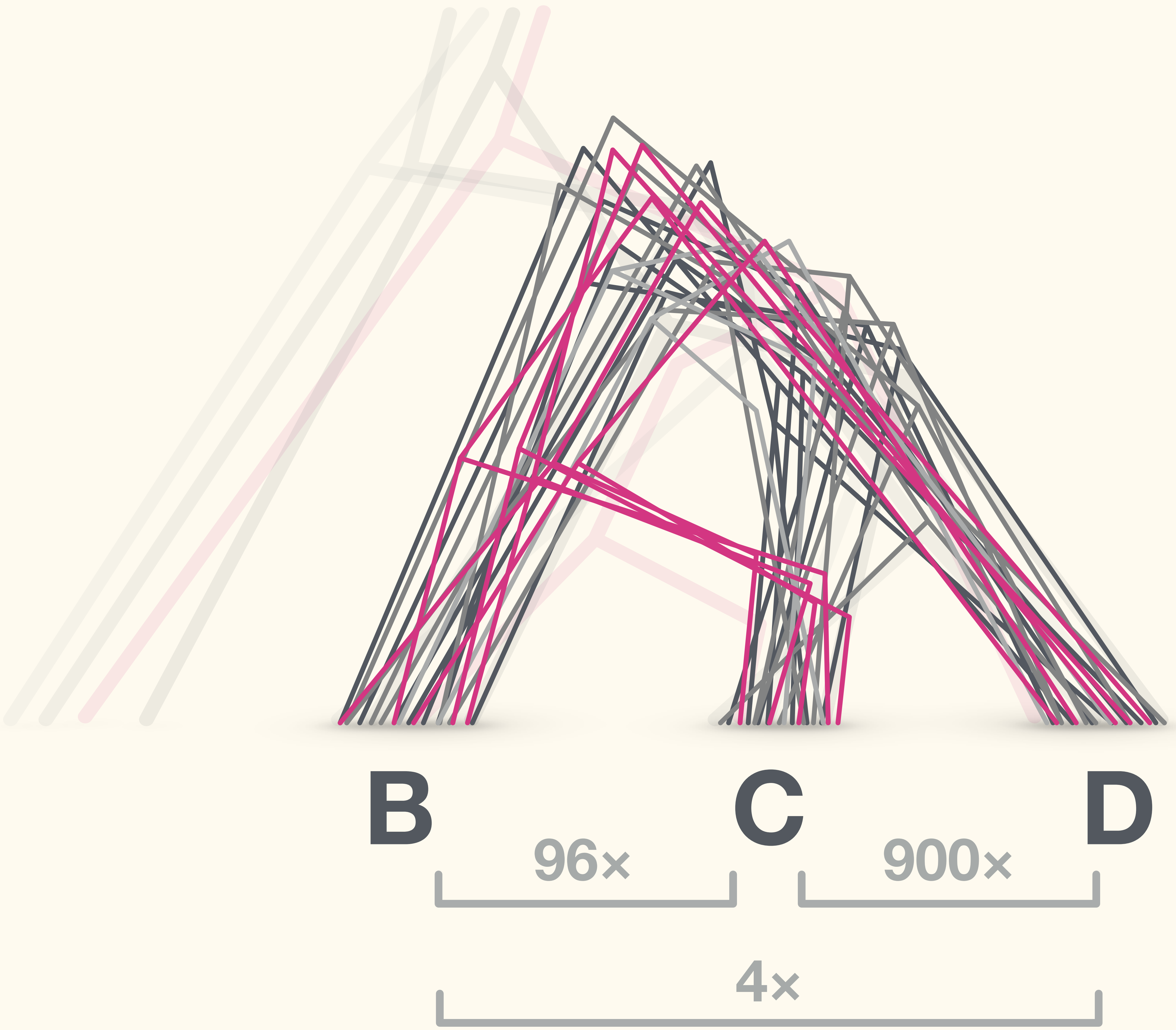
Introgression tests

Gene-tree asymmetry



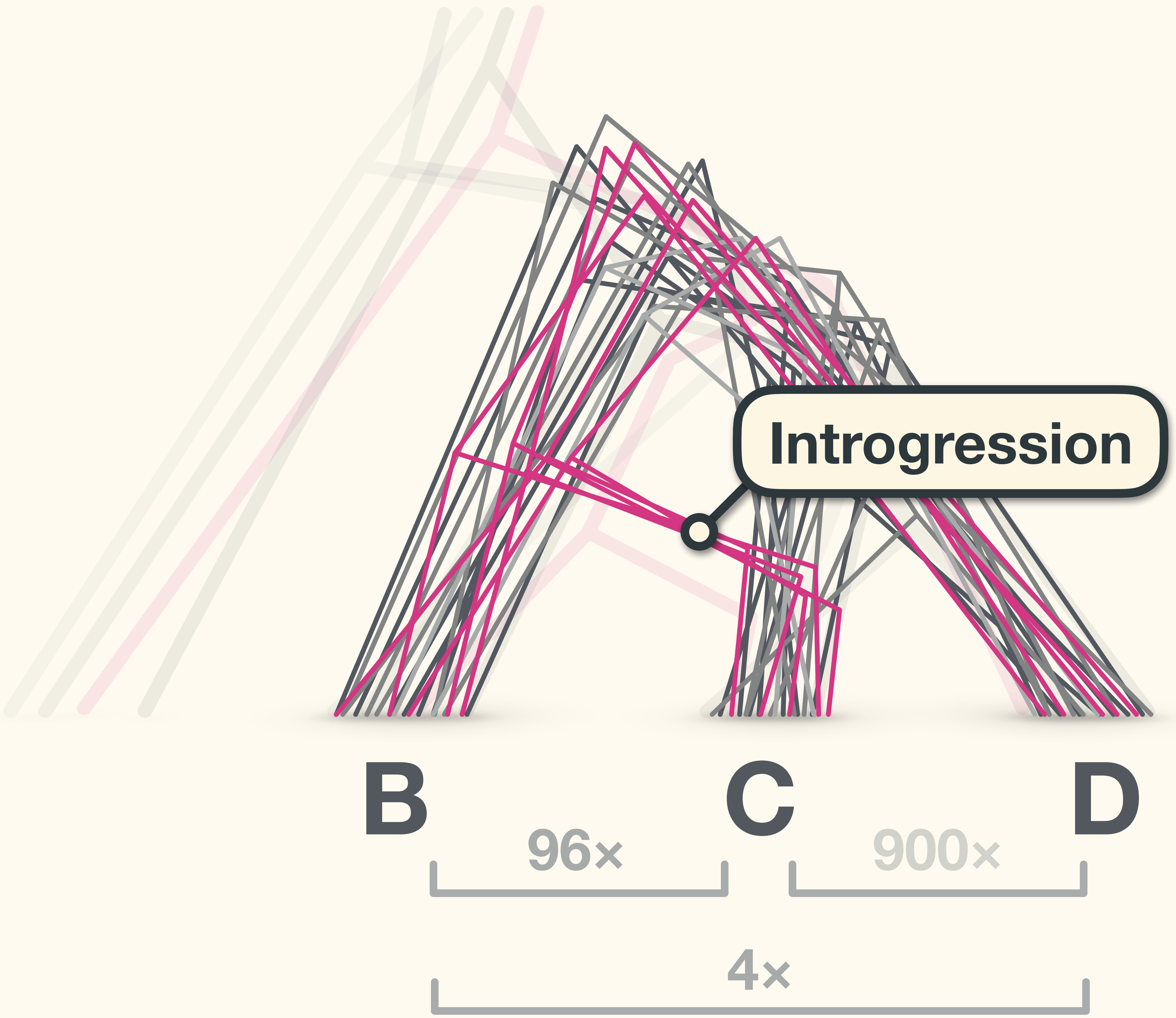
Introgression tests

Gene-tree asymmetry



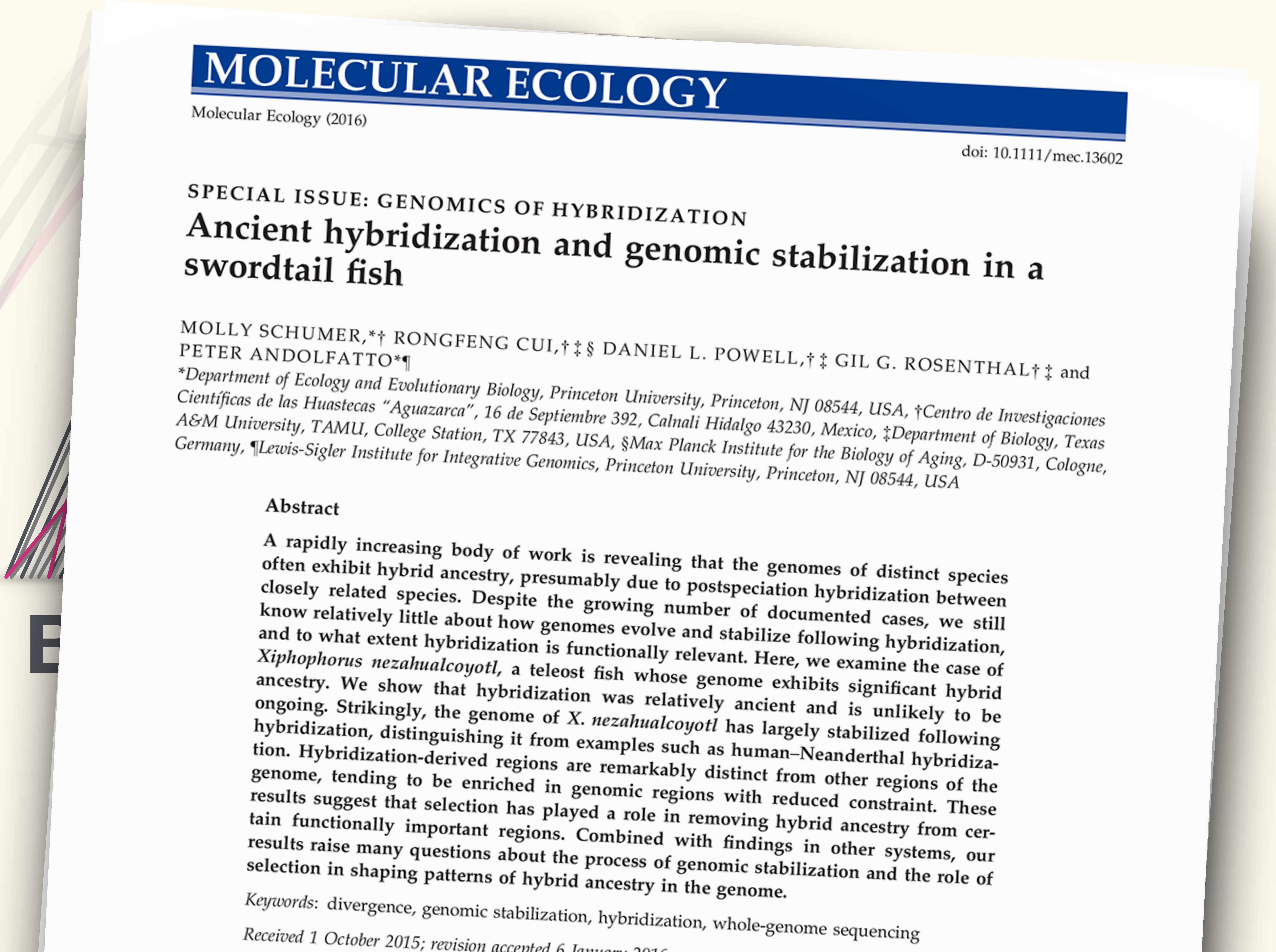
Introgression tests

Gene-tree asymmetry



Introgression tests

Gene-tree asymmetry



MOLECULAR ECOLOGY

Molecular Ecology (2016)

doi: 10.1111/mec.13602

SPECIAL ISSUE: GENOMICS OF HYBRIDIZATION

Ancient hybridization and genomic stabilization in a swordtail fish

MOLLY SCHUMER,*† RONGFENG CUI,†‡§ DANIEL L. POWELL,†‡ GIL G. ROSENTHAL†‡ and PETER ANDOLFATTO*¶

**Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA, †Centro de Investigaciones Científicas de las Huastecas "Aguazarca", 16 de Septiembre 392, Calnali Hidalgo 43230, Mexico, ‡Department of Biology, Texas A&M University, TAMU, College Station, TX 77843, USA, §Max Planck Institute for the Biology of Aging, D-50931, Cologne, Germany, ¶Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA*

Abstract

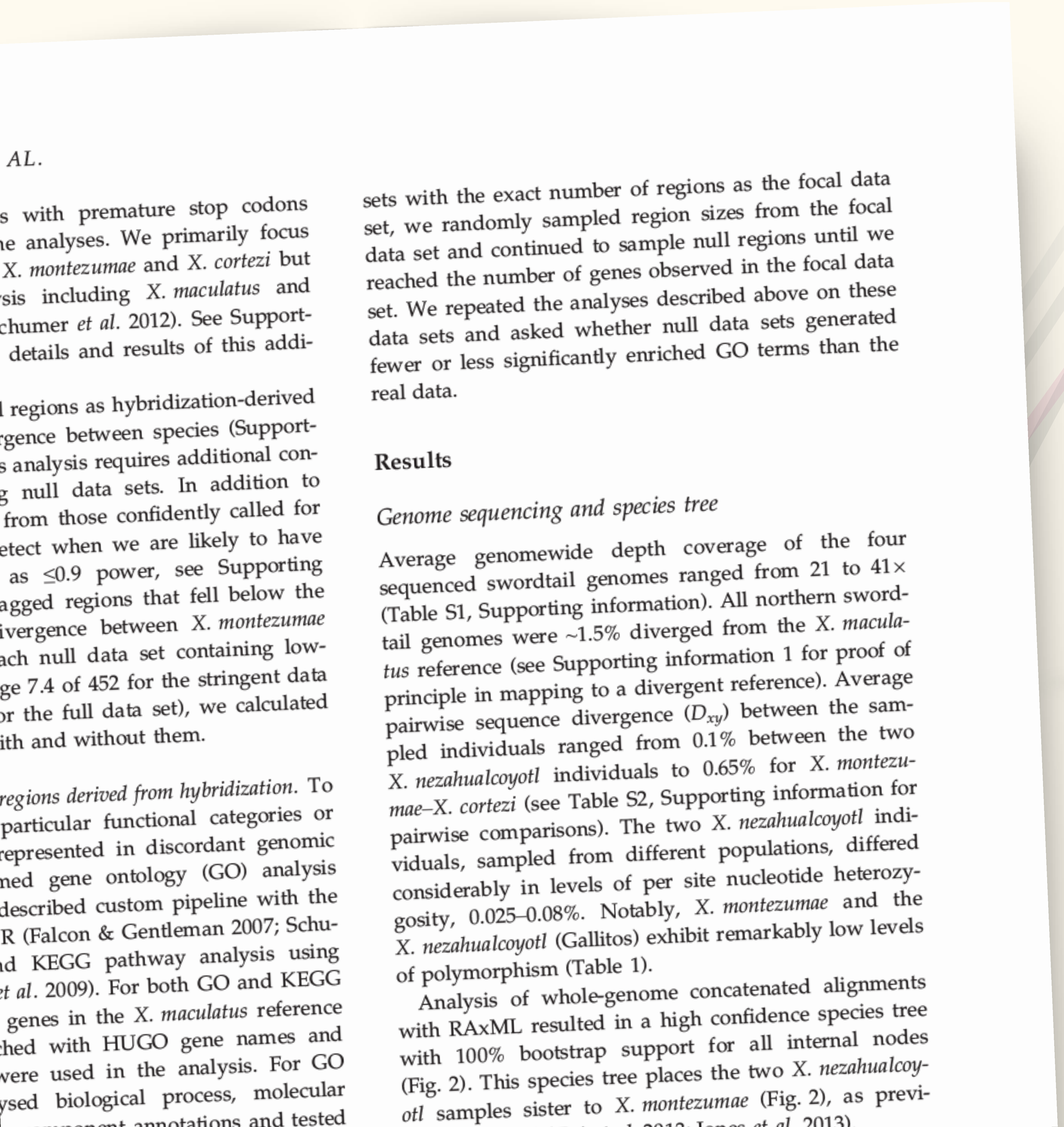
A rapidly increasing body of work is revealing that the genomes of distinct species often exhibit hybrid ancestry, presumably due to postspeciation hybridization between closely related species. Despite the growing number of documented cases, we still know relatively little about how genomes evolve and stabilize following hybridization, and to what extent hybridization is functionally relevant. Here, we examine the case of *Xiphophorus nezahualcoyotl*, a teleost fish whose genome exhibits significant hybrid ancestry. We show that hybridization was relatively ancient and is unlikely to be ongoing. Strikingly, the genome of *X. nezahualcoyotl* has largely stabilized following hybridization, distinguishing it from examples such as human–Neanderthal hybridization. Hybridization-derived regions are remarkably distinct from other regions of the genome, tending to be enriched in genomic regions with reduced constraint. These results suggest that selection has played a role in removing hybrid ancestry from certain functionally important regions. Combined with findings in other systems, our results raise many questions about the process of genomic stabilization and the role of selection in shaping patterns of hybrid ancestry in the genome.

Keywords: divergence, genomic stabilization, hybridization, whole-genome sequencing

Received 1 October 2015; revision accepted 6 January 2016

Introgression tests

Gene-tree asymmetry



AL.

s with premature stop codons he analyses. We primarily focus X. *montezumae* and X. *cortezi* but sis including X. *maculatus* and chumer *et al*. 2012). See Supporting details and results of this addi-

l regions as hybridization-derived rgence between species (Supports analysis requires additional con- g null data sets. In addition to from those confidently called for etect when we are likely to have as ≤0.9 power, see Supporting agged regions that fell below the ivergence between X. *montezumae* ach null data set containing low- ge 7.4 of 452 for the stringent data or the full data set), we calculated ith and without them.

regions derived from hybridization. To particular functional categories or represented in discordant genomic ned gene ontology (GO) analysis described custom pipeline with the R (Falcon & Gentleman 2007; Schu- d KEGG pathway analysis using *et al*. 2009). For both GO and KEGG genes in the X. *maculatus* reference hed with HUGO gene names and were used in the analysis. For GO ysed biological process, molecular mponent annotations and tested

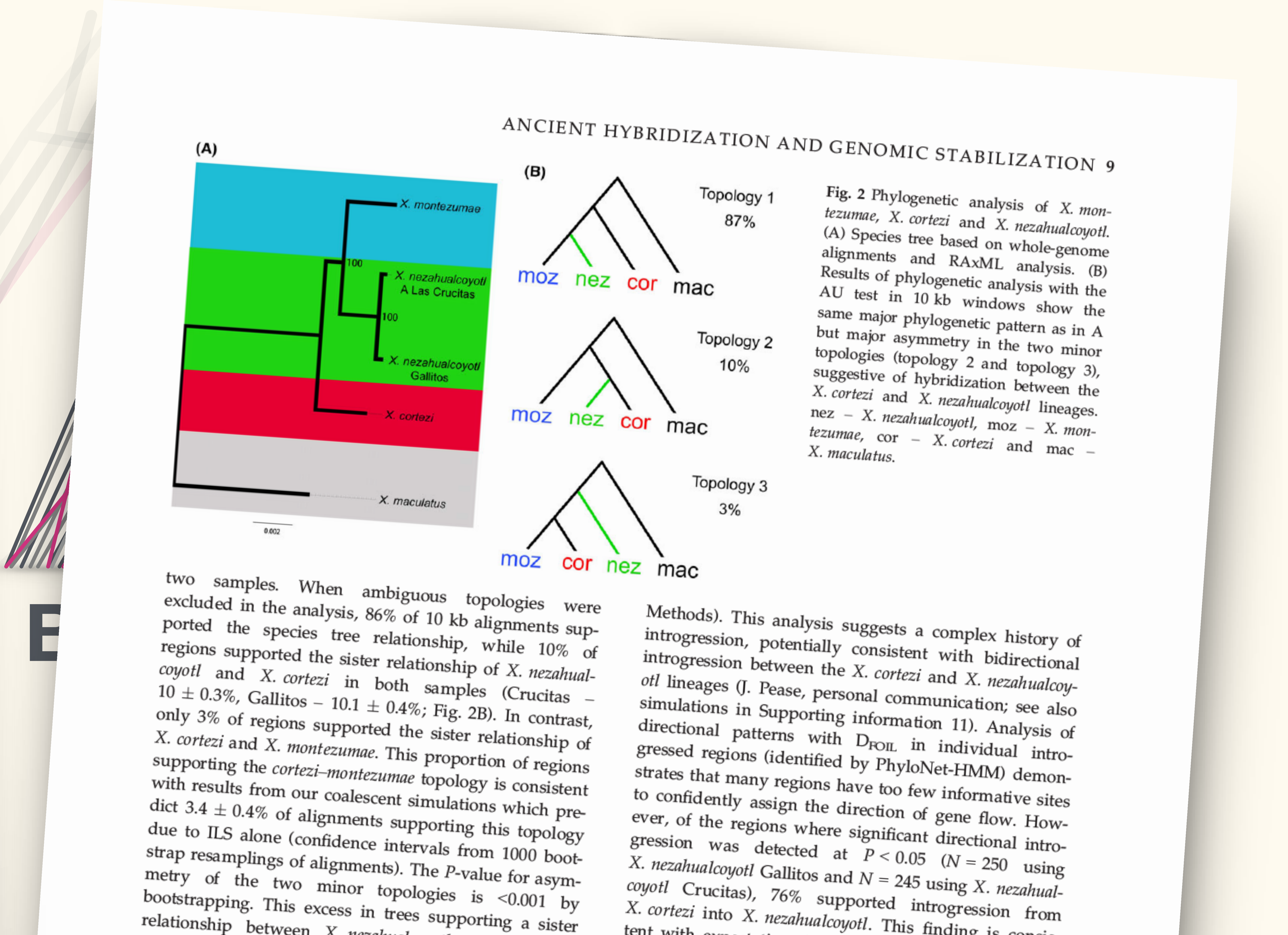
sets with the exact number of regions as the focal data set, we randomly sampled region sizes from the focal data set and continued to sample null regions until we reached the number of genes observed in the focal data set. We repeated the analyses described above on these data sets and asked whether null data sets generated fewer or less significantly enriched GO terms than the real data.

Results

Genome sequencing and species tree

Average genomewide depth coverage of the four sequenced swordtail genomes ranged from 21 to 41× (Table S1, Supporting information). All northern swordtail genomes were ~1.5% diverged from the X. *maculatus* reference (see Supporting information 1 for proof of principle in mapping to a divergent reference). Average pairwise sequence divergence (D_{xy}) between the sampled individuals ranged from 0.1% between the two X. *nezahualcoyotl* individuals to 0.65% for X. *montezumae*–X. *cortezi* (see Table S2, Supporting information for pairwise comparisons). The two X. *nezahualcoyotl* individuals, sampled from different populations, differed considerably in levels of per site nucleotide heterozygosity, 0.025–0.08%. Notably, X. *montezumae* and the X. *nezahualcoyotl* (Gallitos) exhibit remarkably low levels of polymorphism (Table 1).

Analysis of whole-genome concatenated alignments with RAxML resulted in a high confidence species tree with 100% bootstrap support for all internal nodes (Fig. 2). This species tree places the two X. *nezahualcoyotl* samples sister to X. *montezumae* (Fig. 2), as previ-



ANCIENT HYBRIDIZATION AND GENOMIC STABILIZATION 9

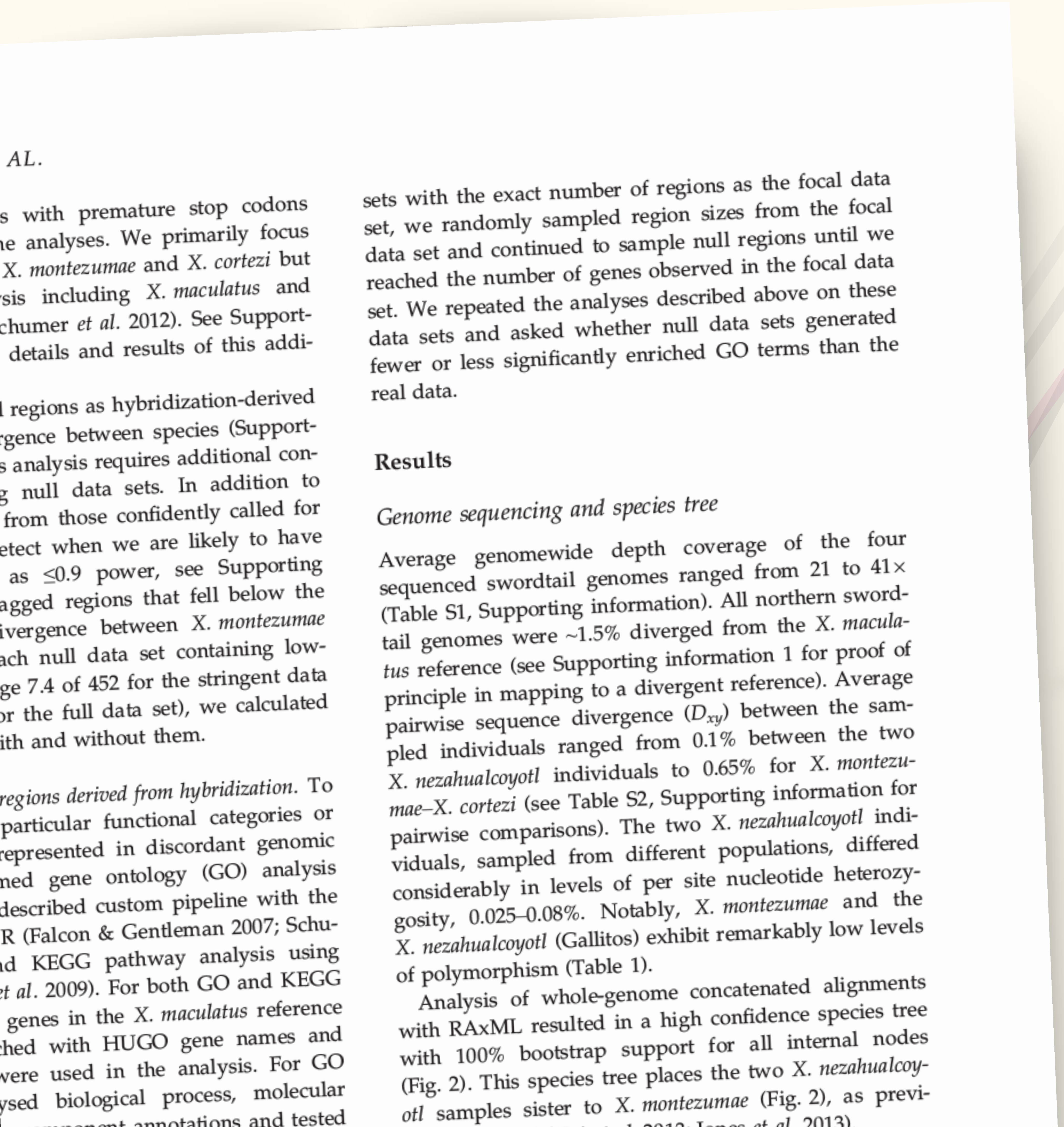
Fig. 2 Phylogenetic analysis of X. *montezumae*, X. *cortezi* and X. *nezahualcoyotl*. (A) Species tree based on whole-genome alignments and RAxML analysis. (B) Results of phylogenetic analysis with the AU test in 10 kb windows show the same major phylogenetic pattern as in A but major asymmetry in the two minor topologies (topology 2 and topology 3), suggestive of hybridization between the X. *cortezi* and X. *nezahualcoyotl* lineages. nez – X. *nezahualcoyotl*, moz – X. *montezumae*, cor – X. *cortezi* and mac – X. *maculatus*.

two samples. When ambiguous topologies were excluded in the analysis, 86% of 10 kb alignments supported the species tree relationship, while 10% of regions supported the sister relationship of X. *nezahualcoyotl* and X. *cortezi* in both samples (Crucitas – 10 ± 0.3%, Gallitos – 10.1 ± 0.4%; Fig. 2B). In contrast, only 3% of regions supported the sister relationship of X. *cortezi* and X. *montezumae*. This proportion of regions supporting the *cortezi–montezumae* topology is consistent with results from our coalescent simulations which predict 3.4 ± 0.4% of alignments supporting this topology due to ILS alone (confidence intervals from 1000 bootstrap resamplings of alignments). The *P*-value for asymmetry of the two minor topologies is <0.001 by bootstrapping. This excess in trees supporting a sister relationship between X. nezahual-

Methods). This analysis suggests a complex history of introgression, potentially consistent with bidirectional introgression between the X. *cortezi* and X. *nezahualcoyotl* lineages (J. Pease, personal communication; see also simulations in Supporting information 11). Analysis of directional patterns with D_{FOIL} in individual introgressed regions (identified by PhyloNet-HMM) demonstrates that many regions have too few informative sites to confidently assign the direction of gene flow. However, of the regions where significant directional introgression was detected at $P < 0.05$ ($N = 250$ using X. *nezahualcoyotl* Gallitos and $N = 245$ using X. *nezahualcoyotl* Crucitas), 76% supported introgression from X. *cortezi* into X. *nezahualcoyotl*. This finding is consis-

Introgression tests

Gene-tree asymmetry



s with premature stop codons he analyses. We primarily focus X. *montezumae* and X. *cortezi* but sis including X. *maculatus* and chumer *et al*. 2012). See Supporting details and results of this addi-

l regions as hybridization-derived rgence between species (Supports analysis requires additional cong null data sets. In addition to from those confidently called for etect when we are likely to have as ≤0.9 power, see Supporting agged regions that fell below the ivergence between X. *montezumae* ach null data set containing lowge 7.4 of 452 for the stringent data or the full data set), we calculated ith and without them.

regions derived from hybridization. To particular functional categories or represented in discordant genomic med gene ontology (GO) analysis described custom pipeline with the R (Falcon & Gentleman 2007; Schund KEGG pathway analysis using *et al*. 2009). For both GO and KEGG genes in the X. *maculatus* reference hed with HUGO gene names and were used in the analysis. For GO ysed biological process, molecular

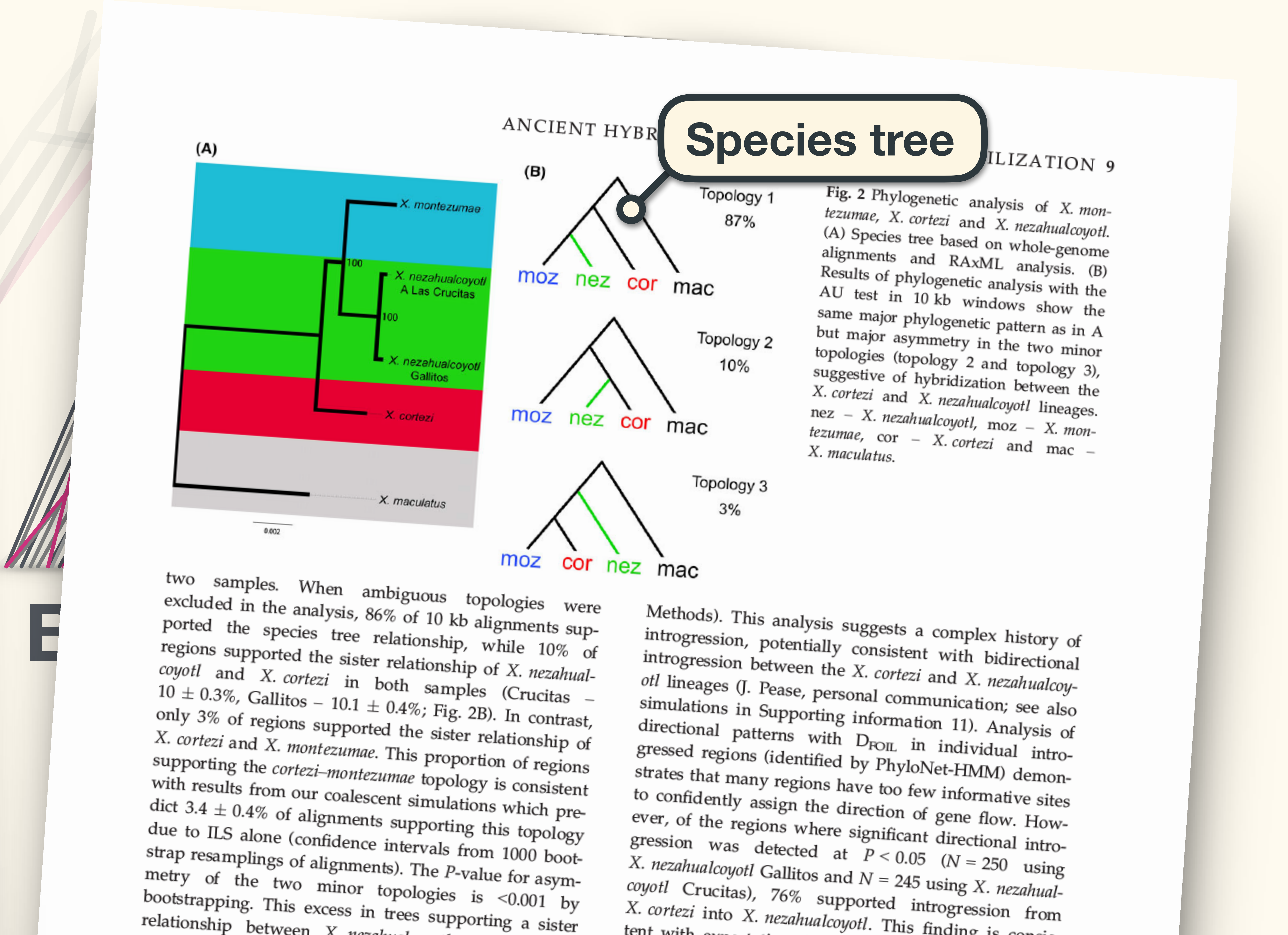
sets with the exact number of regions as the focal data set, we randomly sampled region sizes from the focal data set and continued to sample null regions until we reached the number of genes observed in the focal data set. We repeated the analyses described above on these data sets and asked whether null data sets generated fewer or less significantly enriched GO terms than the real data.

Results

Genome sequencing and species tree

Average genomewide depth coverage of the four sequenced swordtail genomes ranged from 21 to 41× (Table S1, Supporting information). All northern swordtail genomes were ~1.5% diverged from the X. *maculatus* reference (see Supporting information 1 for proof of principle in mapping to a divergent reference). Average pairwise sequence divergence (D_{xy}) between the sampled individuals ranged from 0.1% between the two X. *nezahualcoyotl* individuals to 0.65% for X. *montezumae–*X. *cortezi* (see Table S2, Supporting information for pairwise comparisons). The two X. *nezahualcoyotl* individuals, sampled from different populations, differed considerably in levels of per site nucleotide heterozygosity, 0.025–0.08%. Notably, X. *montezumae* and the X. *nezahualcoyotl* (Gallitos) exhibit remarkably low levels of polymorphism (Table 1).

Analysis of whole-genome concatenated alignments with RAxML resulted in a high confidence species tree with 100% bootstrap support for all internal nodes (Fig. 2). This species tree places the two X. *nezahualcoyotl* samples sister to X. *montezumae* (Fig. 2), as previ-



Species tree

(A)
X. montezumae
100
X. nezahualcoyotl A Las Crucitas
100
X. nezahualcoyotl Gallitos
X. cortezi
X. maculatus
0.002
(B)
Topology 1
87%
moz nez cor mac
Topology 2
10%
moz nez cor mac
Topology 3
3%
moz cor nez mac

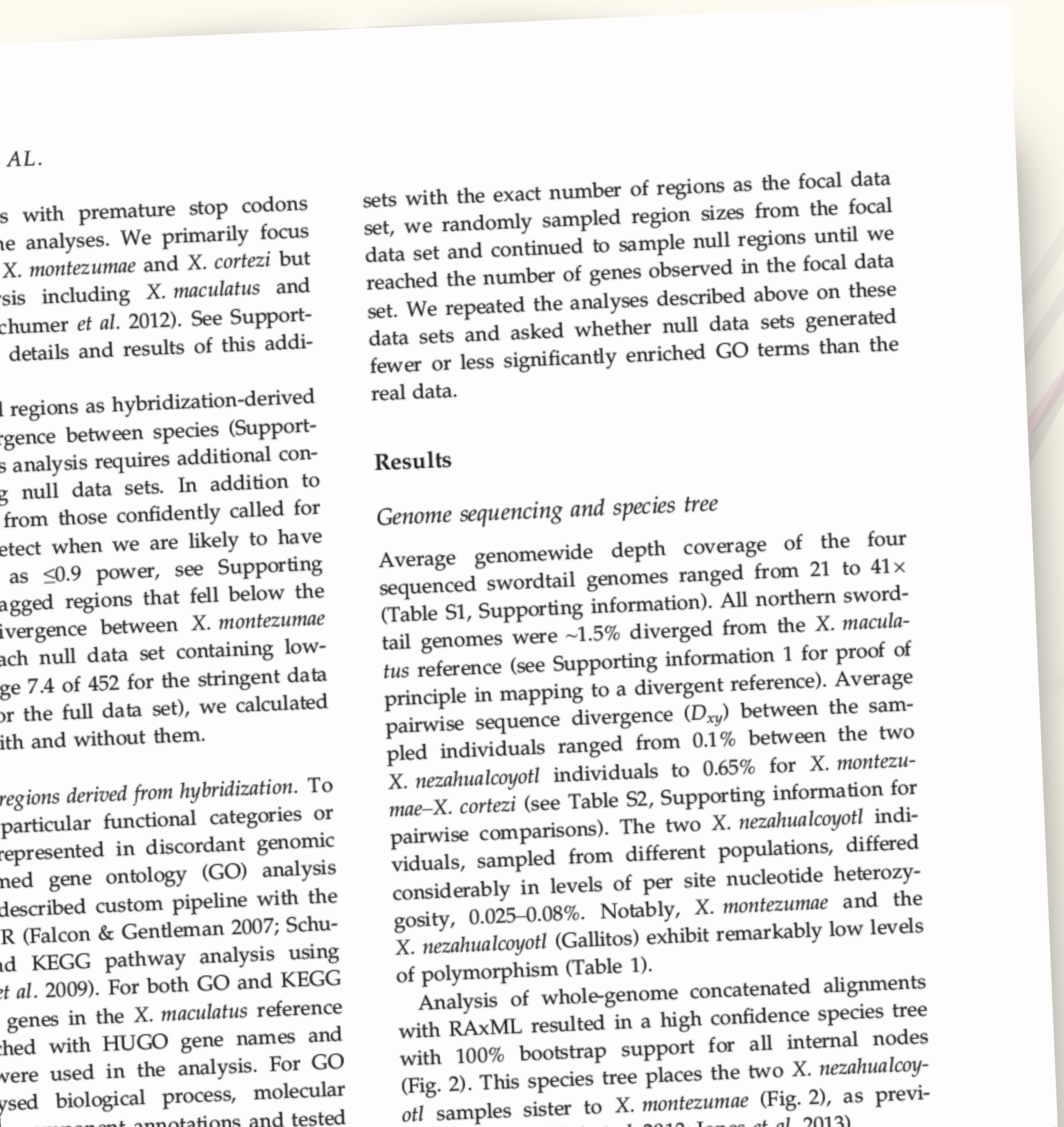
Fig. 2 Phylogenetic analysis of *X. montezumae*, *X. cortezi* and *X. nezahualcoyotl*. (A) Species tree based on whole-genome alignments and RAxML analysis. (B) Results of phylogenetic analysis with the AU test in 10 kb windows show the same major phylogenetic pattern as in A but major asymmetry in the two minor topologies (topology 2 and topology 3), suggestive of hybridization between the *X. cortezi* and *X. nezahualcoytl* lineages. nez – *X. nezahualcoyotl*, moz – *X. montezumae*, cor – *X. cortezi* and mac – *X. maculatus*.

two samples. When ambiguous topologies were excluded in the analysis, 86% of 10 kb alignments supported the species tree relationship, while 10% of regions supported the sister relationship of *X. nezahualcoyotl* and *X. cortezi* in both samples (Crucitas – 10 ± 0.3%, Gallitos – 10.1 ± 0.4%; Fig. 2B). In contrast, only 3% of regions supported the sister relationship of *X. cortezi* and *X. montezumae*. This proportion of regions supporting the *cortezi–montezumae* topology is consistent with results from our coalescent simulations which predict 3.4 ± 0.4% of alignments supporting this topology due to ILS alone (confidence intervals from 1000 bootstrap resamplings of alignments). The *P*-value for asymmetry of the two minor topologies is <0.001 by bootstrapping. This excess in trees supporting a sister relationship between

Methods). This analysis suggests a complex history of introgression, potentially consistent with bidirectional introgression between the *X. cortezi* and *X. nezahualcoyotl* lineages (J. Pease, personal communication; see also simulations in Supporting information 11). Analysis of directional patterns with D_{FOIL} in individual introgressed regions (identified by PhyloNet-HMM) demonstrates that many regions have too few informative sites to confidently assign the direction of gene flow. However, of the regions where significant directional introgression was detected at $P < 0.05$ ($N = 250$ using *X. nezahualcoyotl* Gallitos and $N = 245$ using *X. nezahualcoyotl* Crucitas), 76% supported introgression from *X. cortezi* into *X. nezahualcoyotl*. This finding is

Introgression tests

Gene-tree asymmetry



s with premature stop codons
ne analyses. We primarily focus
X. *montezumae* and X. *cortezi* but
sis including X. *maculatus* and
chumer *et al*. 2012). See Support-
details and results of this addi-

l regions as hybridization-derived
rgence between species (Support-
s analysis requires additional con-
g null data sets. In addition to
from those confidently called for
etect when we are likely to have
as ≤0.9 power, see Supporting
agged regions that fell below the
ivergence between X. *montezumae*
ach null data set containing low-
ge 7.4 of 452 for the stringent data
or the full data set), we calculated
ith and without them.

regions derived from hybridization. To
particular functional categories or
represented in discordant genomic
ned gene ontology (GO) analysis
described custom pipeline with the
R (Falcon & Gentleman 2007; Schu-
nd KEGG pathway analysis using
et al. 2009). For both GO and KEGG
genes in the X. *maculatus* reference
ched with HUGO gene names and
were used in the analysis. For GO
ysed biological process, molecular

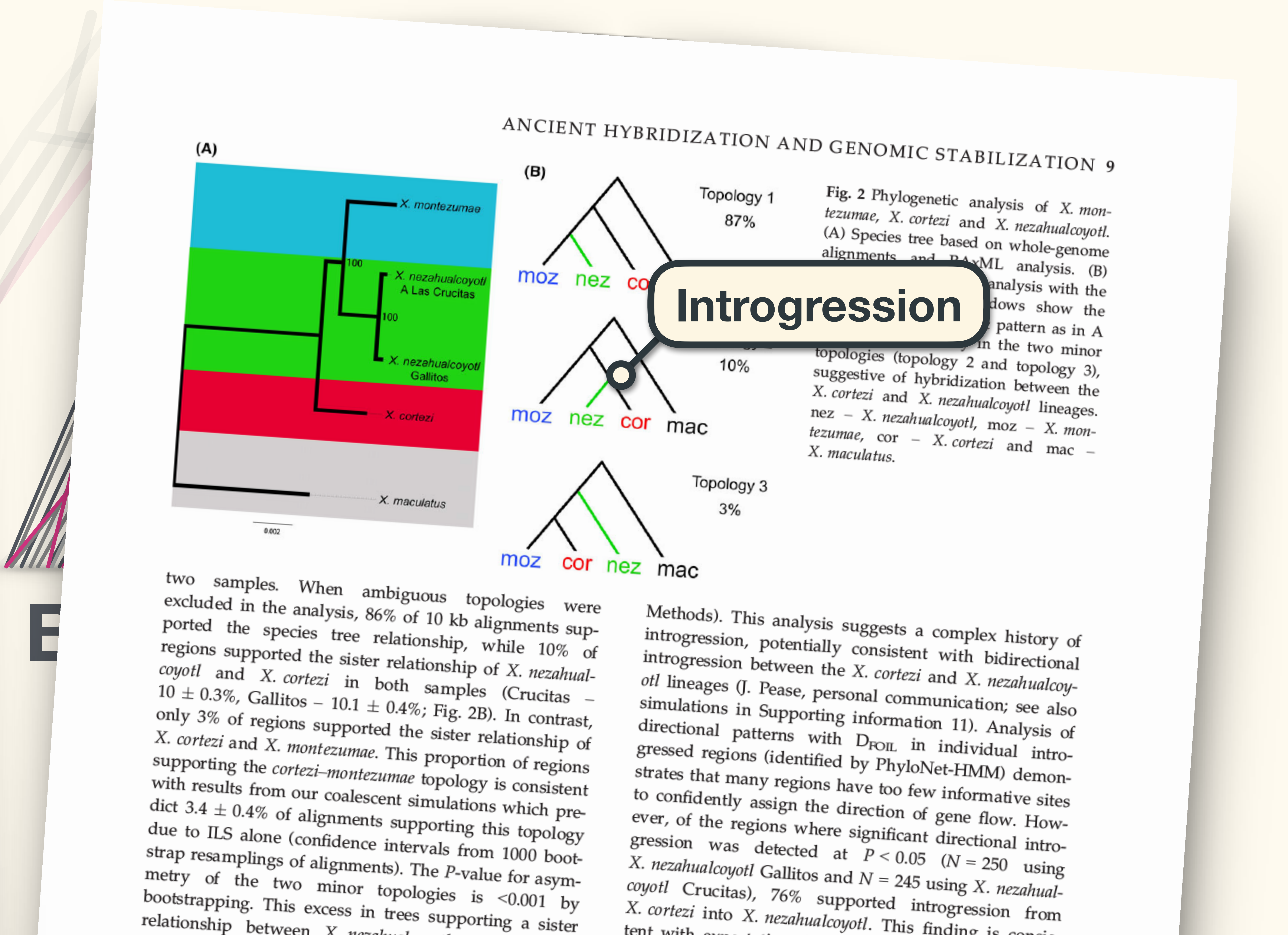
sets with the exact number of regions as the focal data set, we randomly sampled region sizes from the focal data set and continued to sample null regions until we reached the number of genes observed in the focal data set. We repeated the analyses described above on these data sets and asked whether null data sets generated fewer or less significantly enriched GO terms than the real data.

Results

Genome sequencing and species tree

Average genomewide depth coverage of the four sequenced swordtail genomes ranged from 21 to 41× (Table S1, Supporting information). All northern swordtail genomes were ~1.5% diverged from the X. *maculatus* reference (see Supporting information 1 for proof of principle in mapping to a divergent reference). Average pairwise sequence divergence (D_{xy}) between the sampled individuals ranged from 0.1% between the two X. *nezahualcoyotl* individuals to 0.65% for X. *montezumae–*X. *cortezi* (see Table S2, Supporting information for pairwise comparisons). The two X. *nezahualcoyotl* individuals, sampled from different populations, differed considerably in levels of per site nucleotide heterozygosity, 0.025–0.08%. Notably, X. *montezumae* and the X. *nezahualcoyotl* (Gallitos) exhibit remarkably low levels of polymorphism (Table 1).

Analysis of whole-genome concatenated alignments with RAxML resulted in a high confidence species tree with 100% bootstrap support for all internal nodes (Fig. 2). This species tree places the two X. *nezahualcoyotl* samples sister to X. *montezumae* (Fig. 2), as previ-



Introgression

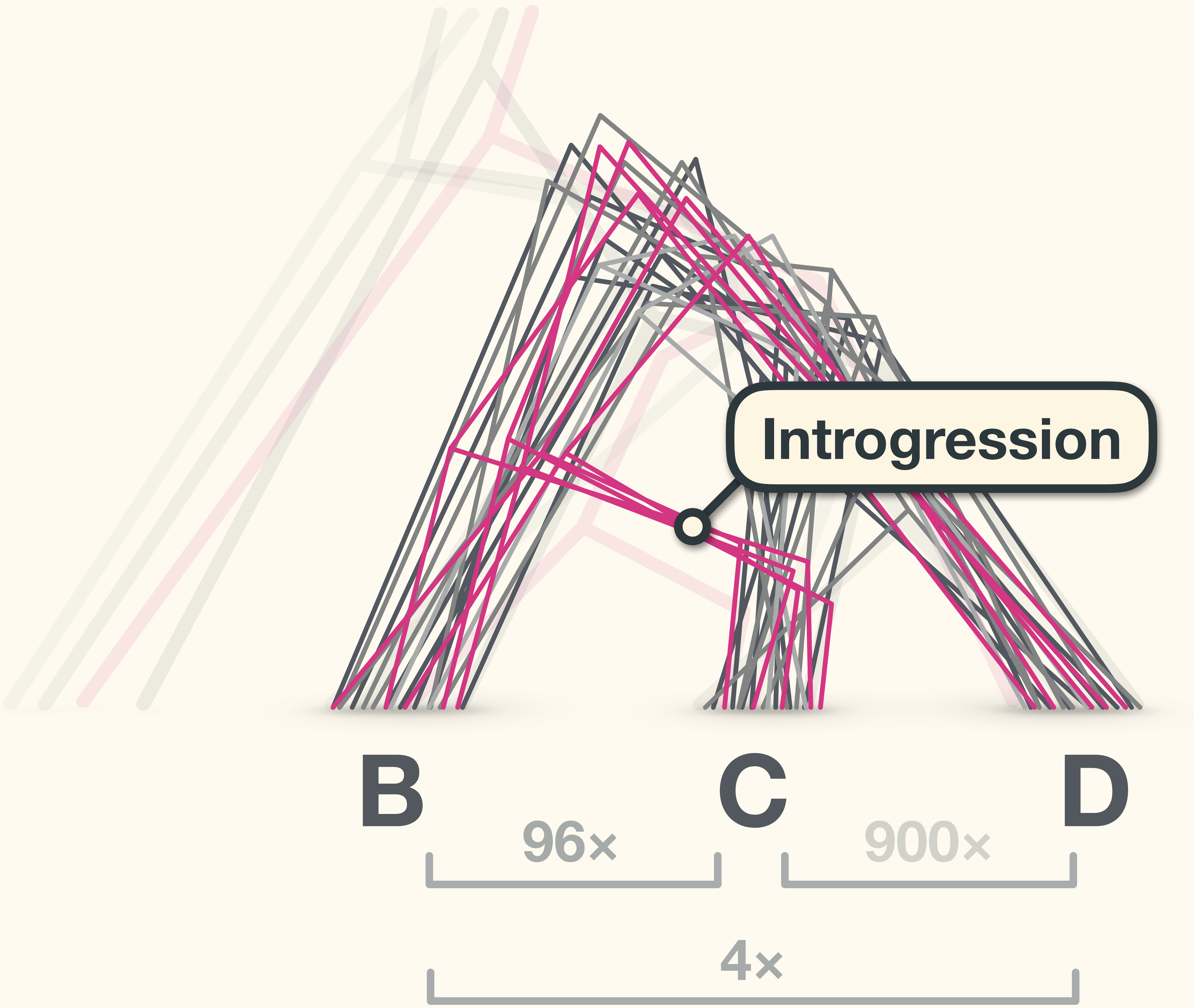
Fig. 2 Phylogenetic analysis of *X. montezumae*, *X. cortezi* and *X. nezahualcoyotl*. (A) Species tree based on whole-genome alignments and RAxML analysis. (B) analysis with the dows show the pattern as in A in the two minor topologies (topology 2 and topology 3), suggestive of hybridization between the *X. cortezi* and *X. nezahualcoyotl* lineages. nez – *X. nezahualcoyotl*, moz – *X. montezumae*, cor – *X. cortezi* and mac – *X. maculatus*.

two samples. When ambiguous topologies were excluded in the analysis, 86% of 10 kb alignments supported the species tree relationship, while 10% of regions supported the sister relationship of X. *nezahualcoyotl* and X. *cortezi* in both samples (Crucitas – 10 ± 0.3%, Gallitos – 10.1 ± 0.4%; Fig. 2B). In contrast, only 3% of regions supported the sister relationship of X. *cortezi* and X. *montezumae*. This proportion of regions supporting the *cortezi–montezumae* topology is consistent with results from our coalescent simulations which predict 3.4 ± 0.4% of alignments supporting this topology due to ILS alone (confidence intervals from 1000 bootstrap resamplings of alignments). The *P*-value for asymmetry of the two minor topologies is <0.001 by bootstrapping. This excess in trees supporting a sister relationship between

Methods). This analysis suggests a complex history of introgression, potentially consistent with bidirectional introgression between the X. *cortezi* and X. *nezahualcoyotl* lineages (J. Pease, personal communication; see also simulations in Supporting information 11). Analysis of directional patterns with D_{FOIL} in individual introgressed regions (identified by PhyloNet-HMM) demonstrates that many regions have too few informative sites to confidently assign the direction of gene flow. However, of the regions where significant directional introgression was detected at $P < 0.05$ ($N = 250$ using X. *nezahualcoyotl* Gallitos and $N = 245$ using X. *nezahualcoyotl* Crucitas), 76% supported introgression from X. *cortezi* into X. *nezahualcoyotl*. This finding is consis-

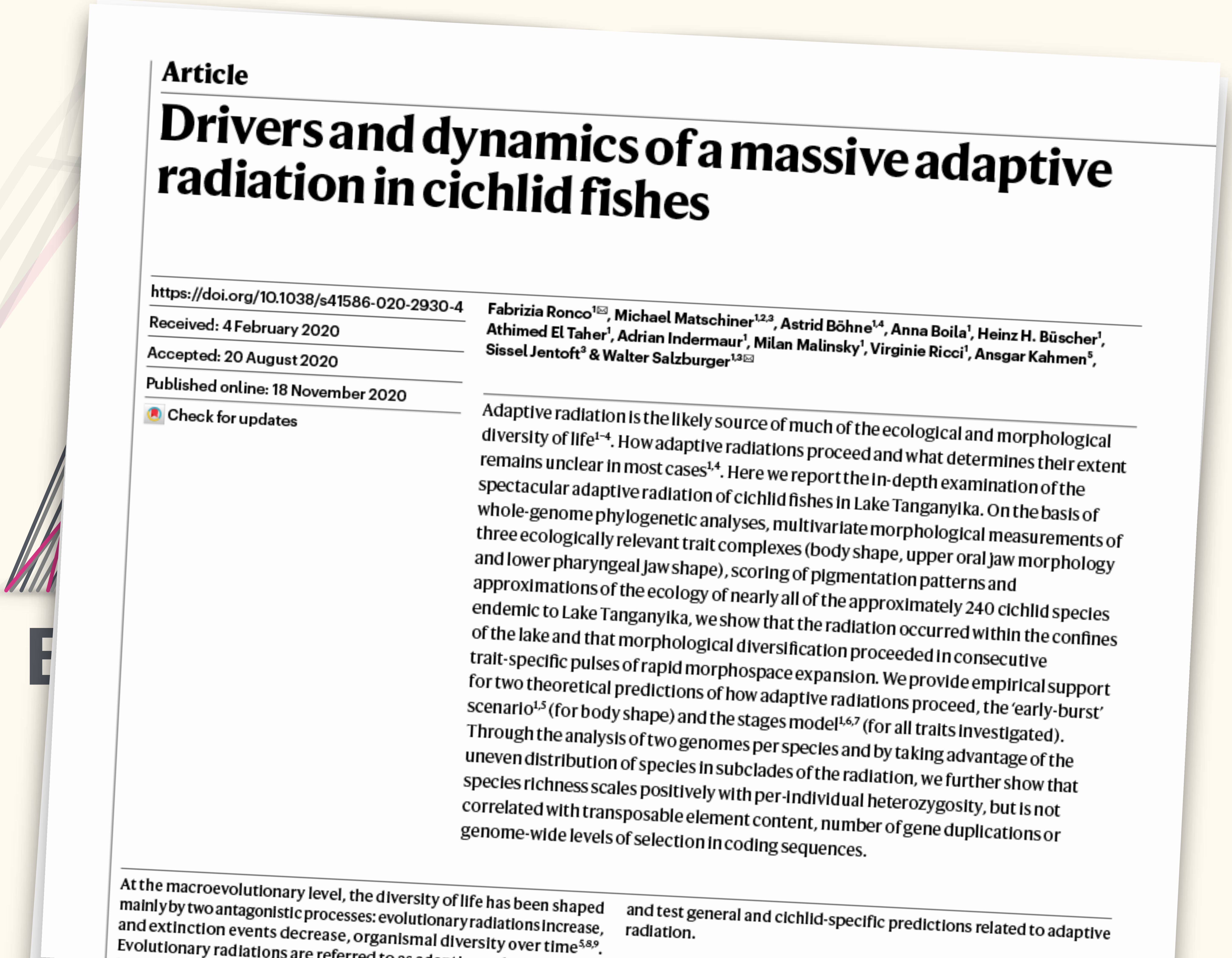
Introgression tests

Gene-tree asymmetry



Introgression tests

Gene-tree asymmetry



Article

Drivers and dynamics of a massive adaptive radiation in cichlid fishes

https://doi.org/10.1038/s41586-020-2930-4

Received: 4 February 2020

Accepted: 20 August 2020

Published online: 18 November 2020

Check for updates

Fabrizia Ronco[1✉], Michael Matschiner[1,2,3], Astrid Böhne[1,4], Anna Boila[1], Heinz H. Büscher[1], Athimed El Taher[1], Adrian Indermaur[1], Milan Malinsky[1], Virginie Ricci[1], Ansgar Kahmen[5], Sissel Jentoft[3] & Walter Salzburger[1,3✉]

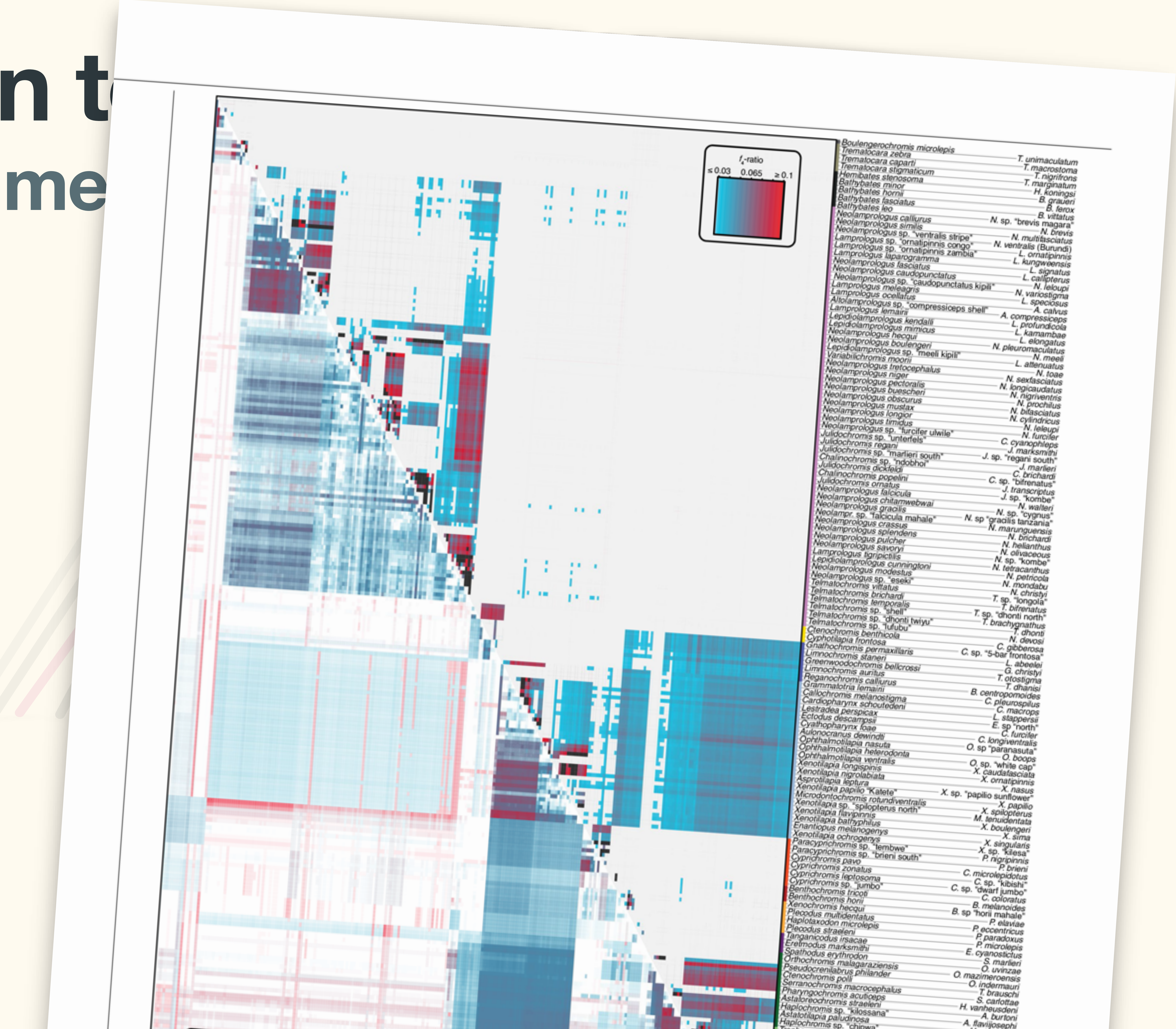
Adaptive radiation is the likely source of much of the ecological and morphological diversity of life[1–4]. How adaptive radiations proceed and what determines their extent remains unclear in most cases[1,4]. Here we report the in-depth examination of the spectacular adaptive radiation of cichlid fishes in Lake Tanganyika. On the basis of whole-genome phylogenetic analyses, multivariate morphological measurements of three ecologically relevant trait complexes (body shape, upper oral jaw morphology and lower pharyngeal jaw shape), scoring of pigmentation patterns and approximations of the ecology of nearly all of the approximately 240 cichlid species endemic to Lake Tanganyika, we show that the radiation occurred within the confines of the lake and that morphological diversification proceeded in consecutive trait-specific pulses of rapid morphospace expansion. We provide empirical support for two theoretical predictions of how adaptive radiations proceed, the 'early-burst' scenario[1,5] (for body shape) and the stages model[1,6,7] (for all traits investigated). Through the analysis of two genomes per species and by taking advantage of the uneven distribution of species in subclades of the radiation, we further show that species richness scales positively with per-individual heterozygosity, but is not correlated with transposable element content, number of gene duplications or genome-wide levels of selection in coding sequences.

At the macroevolutionary level, the diversity of life has been shaped mainly by two antagonistic processes: evolutionary radiations increase, and extinction events decrease, organismal diversity over time[5,8,9]. Evolutionary radiations are referred to

and test general and cichlid-specific predictions related to adaptive radiation.

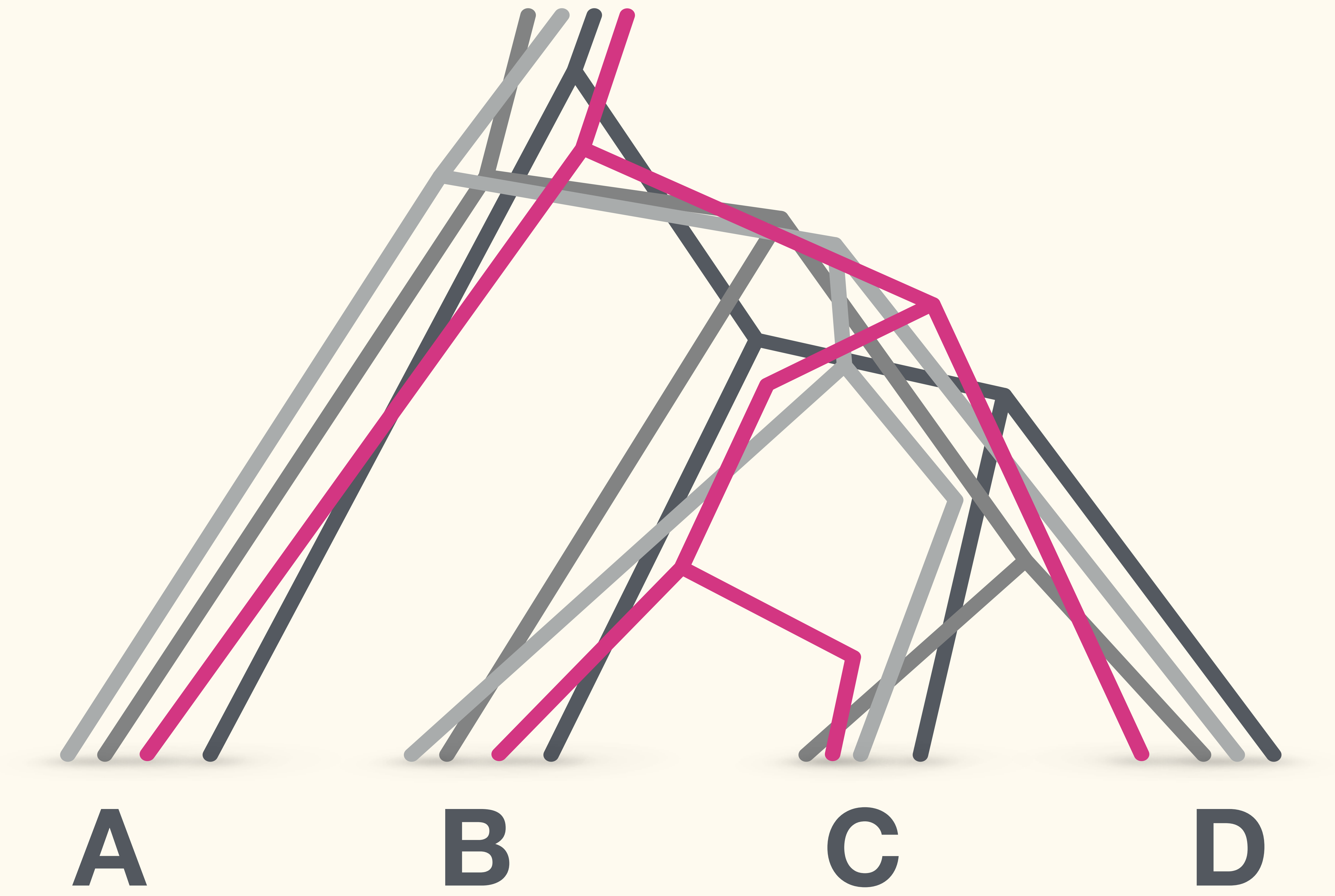
Introgression t

Gene-tree asymme



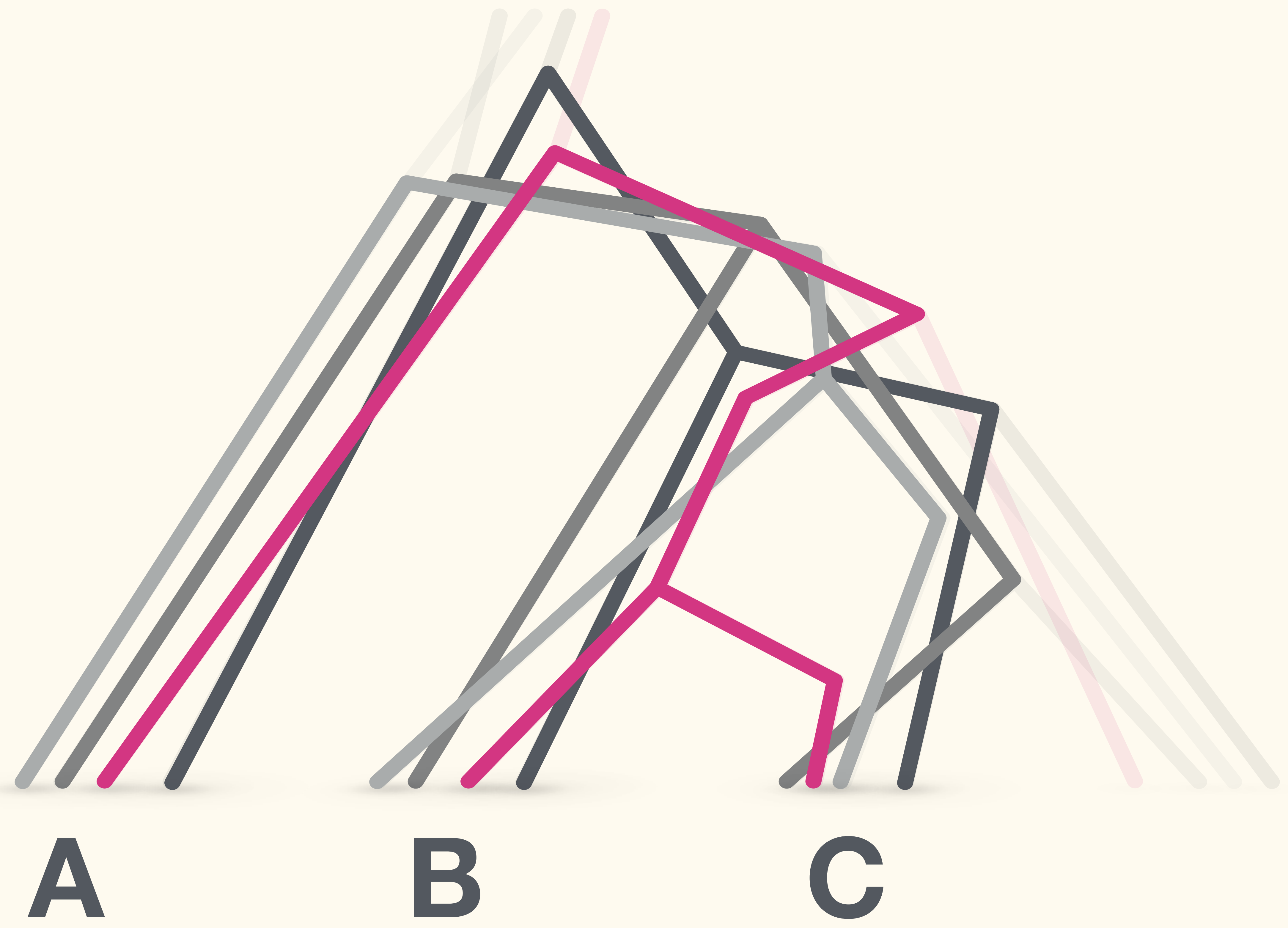
Introgression tests

Divergence times



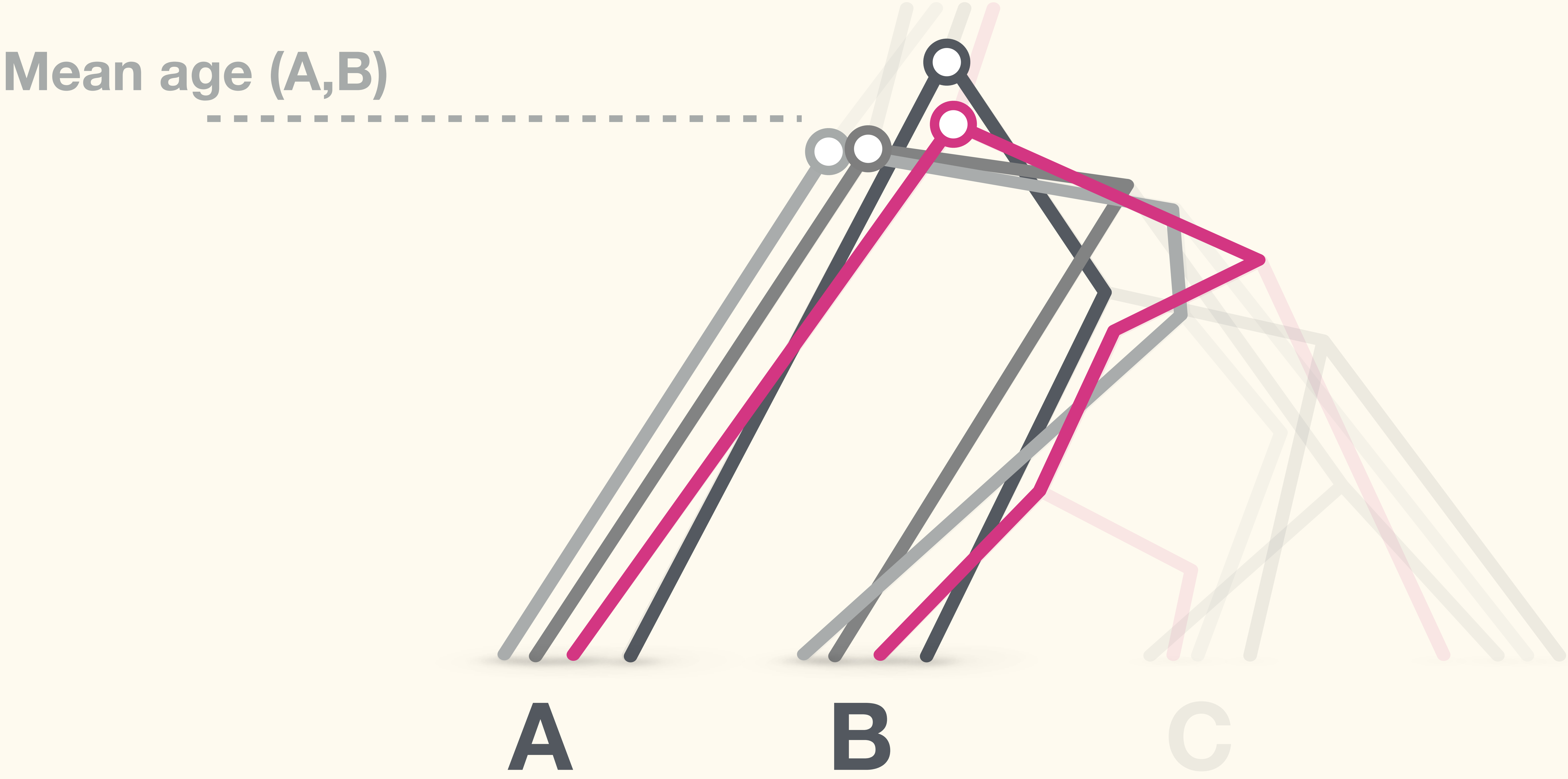
Introgression tests

Divergence times



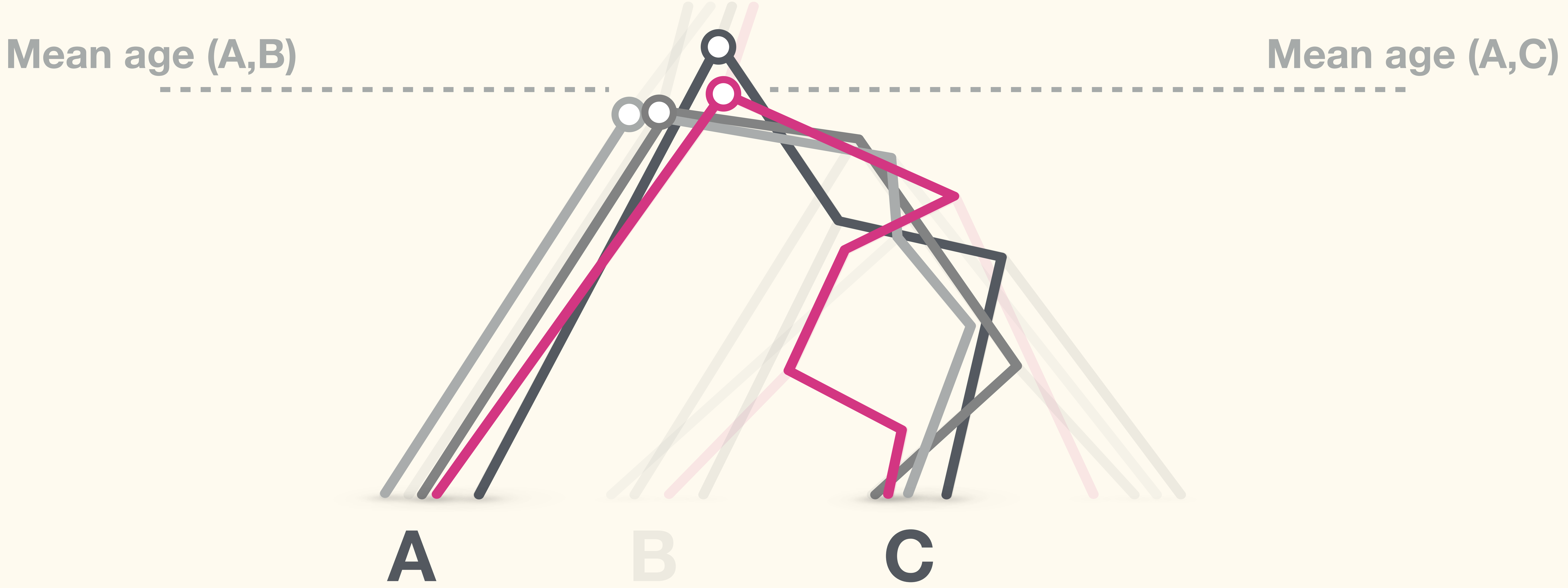
Introgression tests

Divergence times



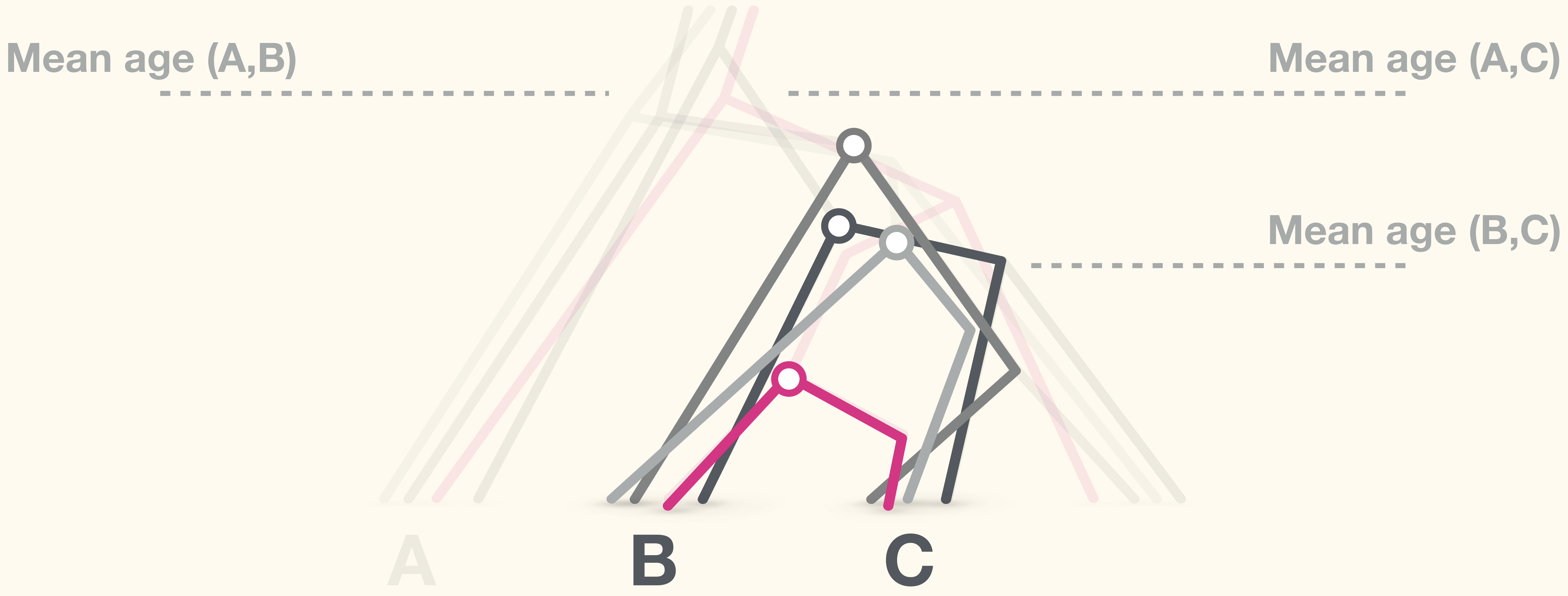
Introgression tests

Divergence times



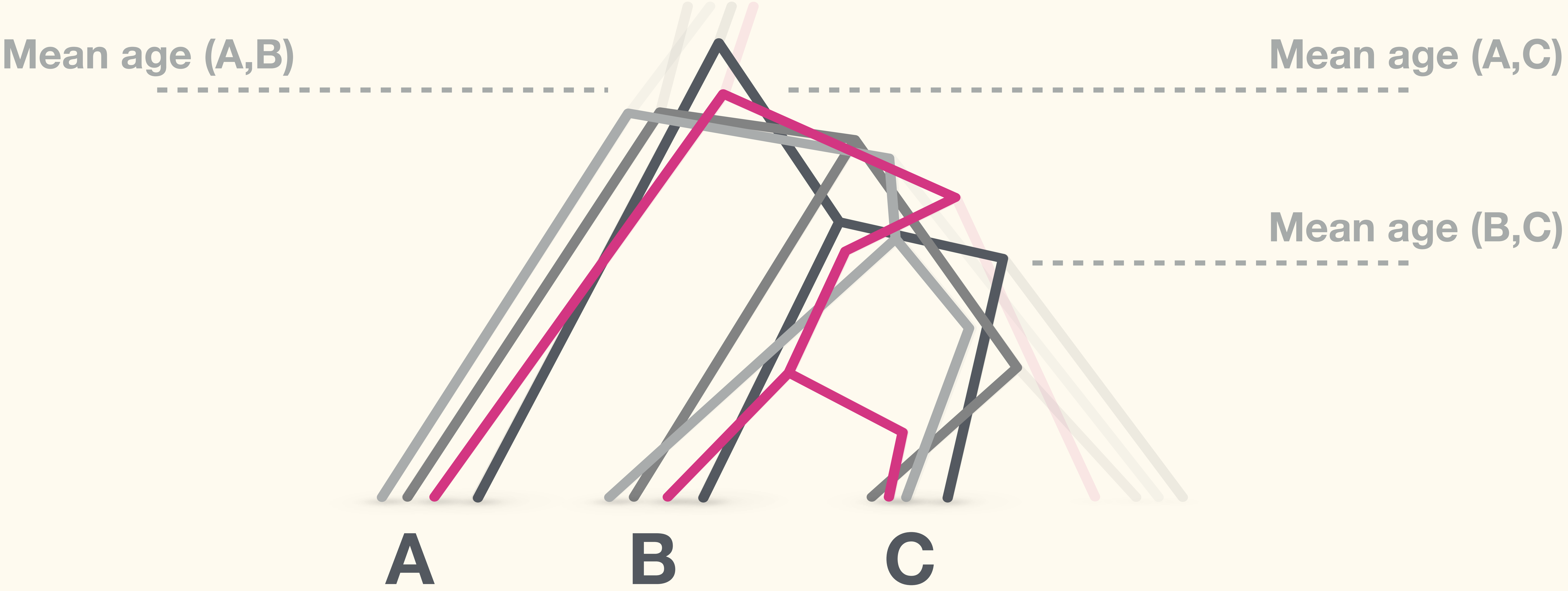
Introgression tests

Divergence times



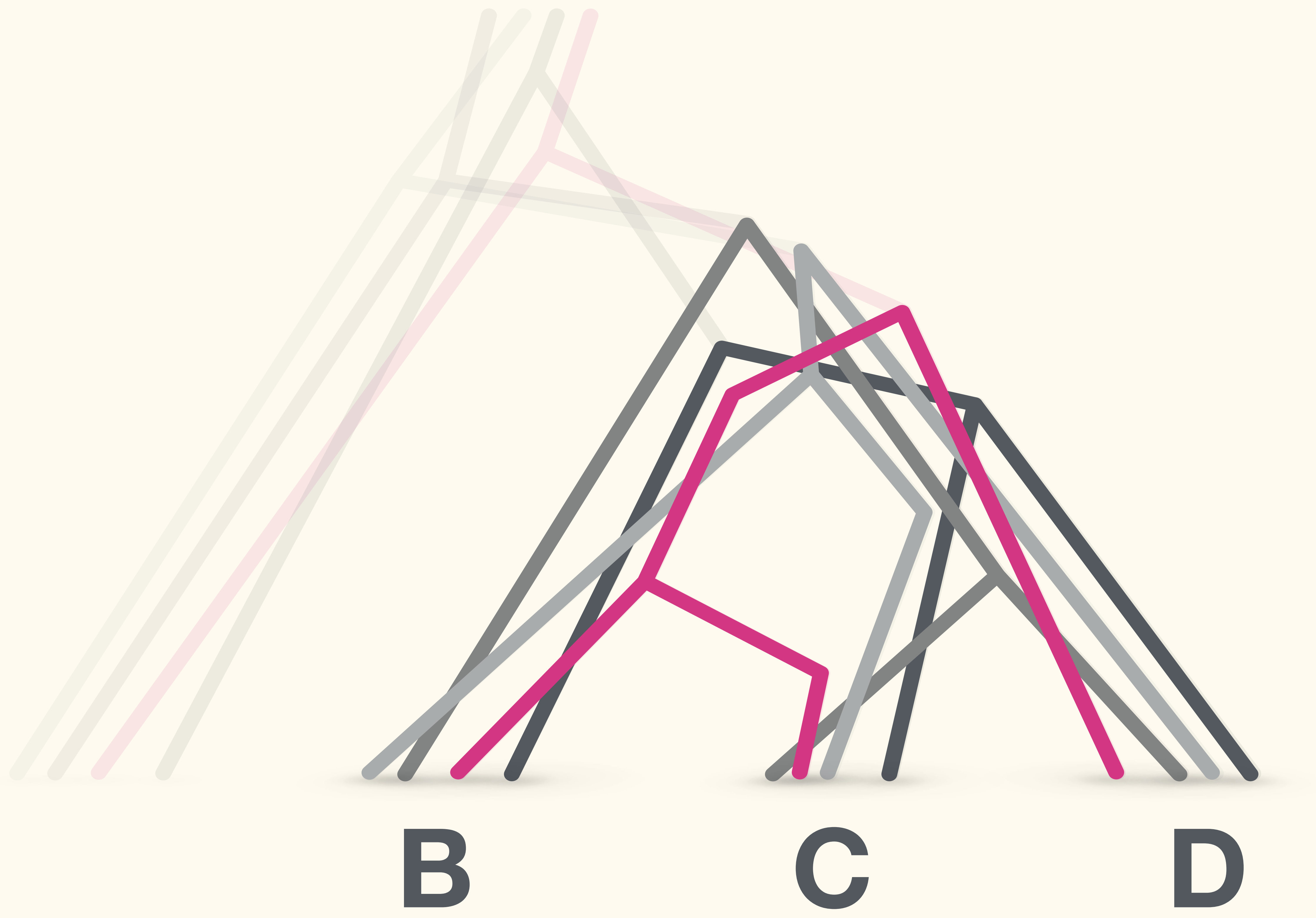
Introgression tests

Divergence times



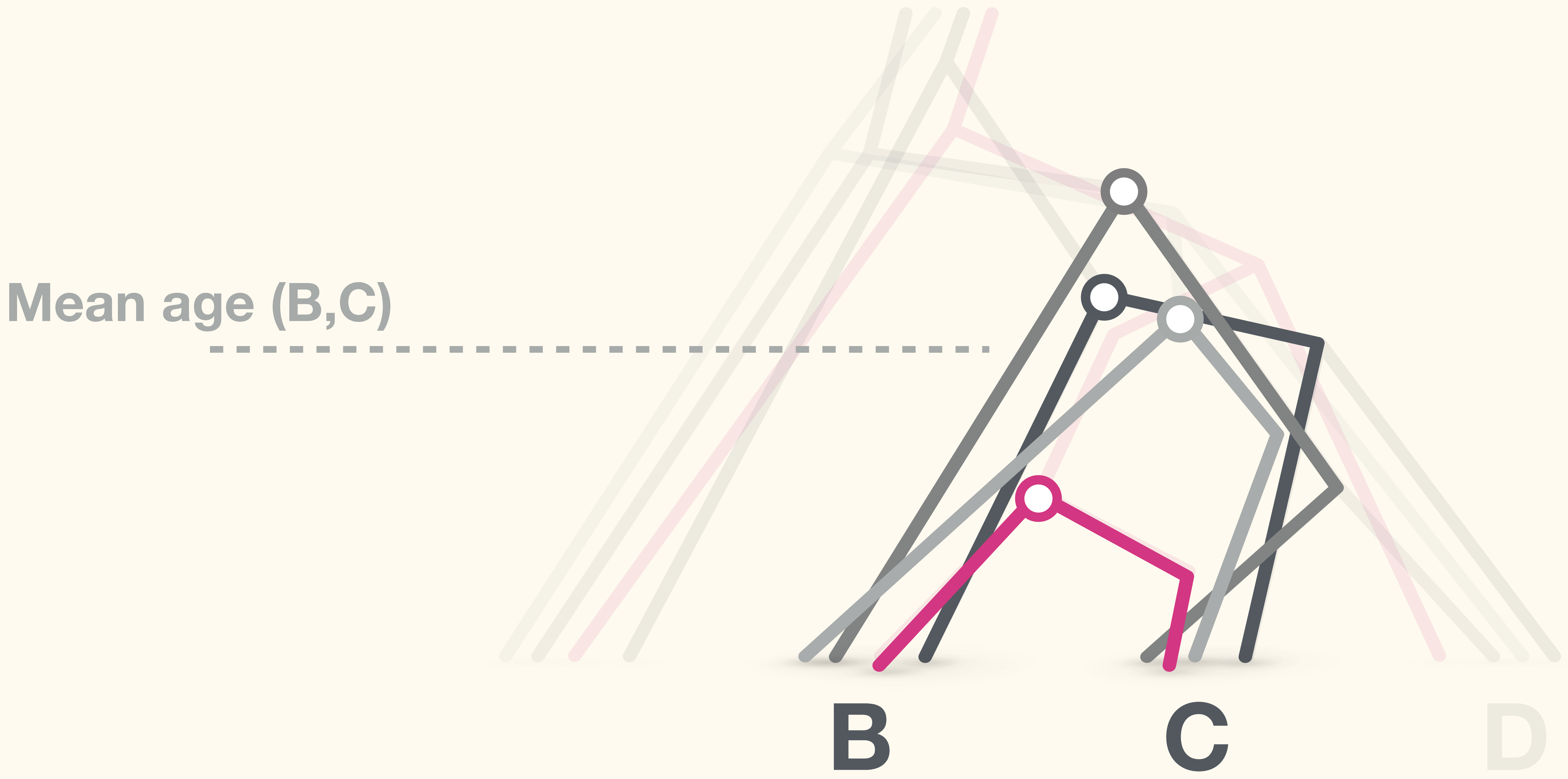
Introgression tests

Divergence times



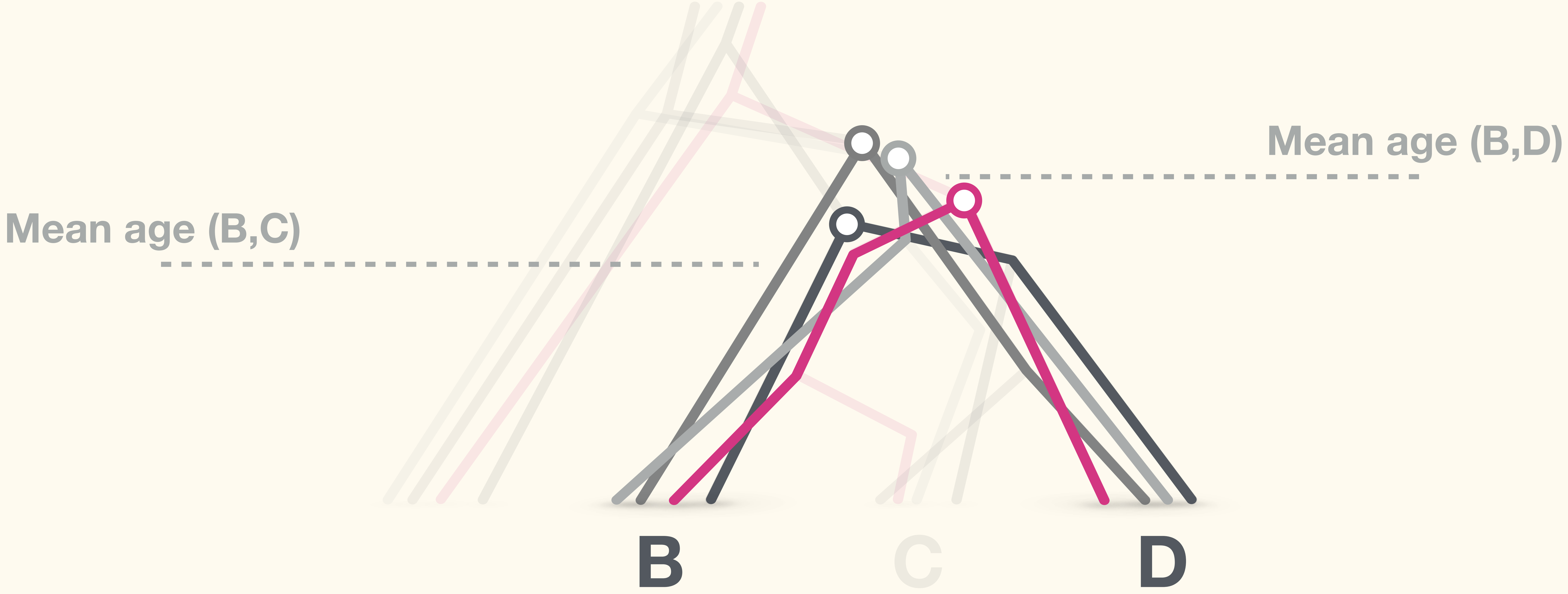
Introgression tests

Divergence times



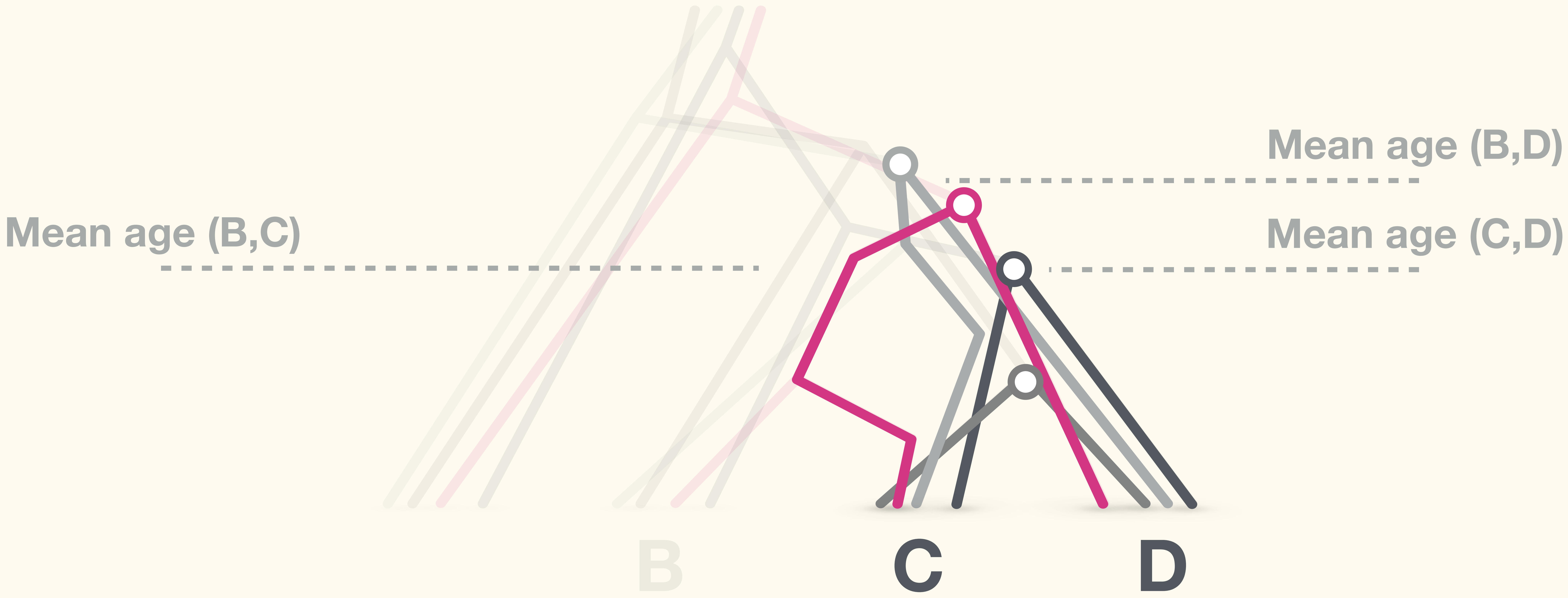
Introgression tests

Divergence times



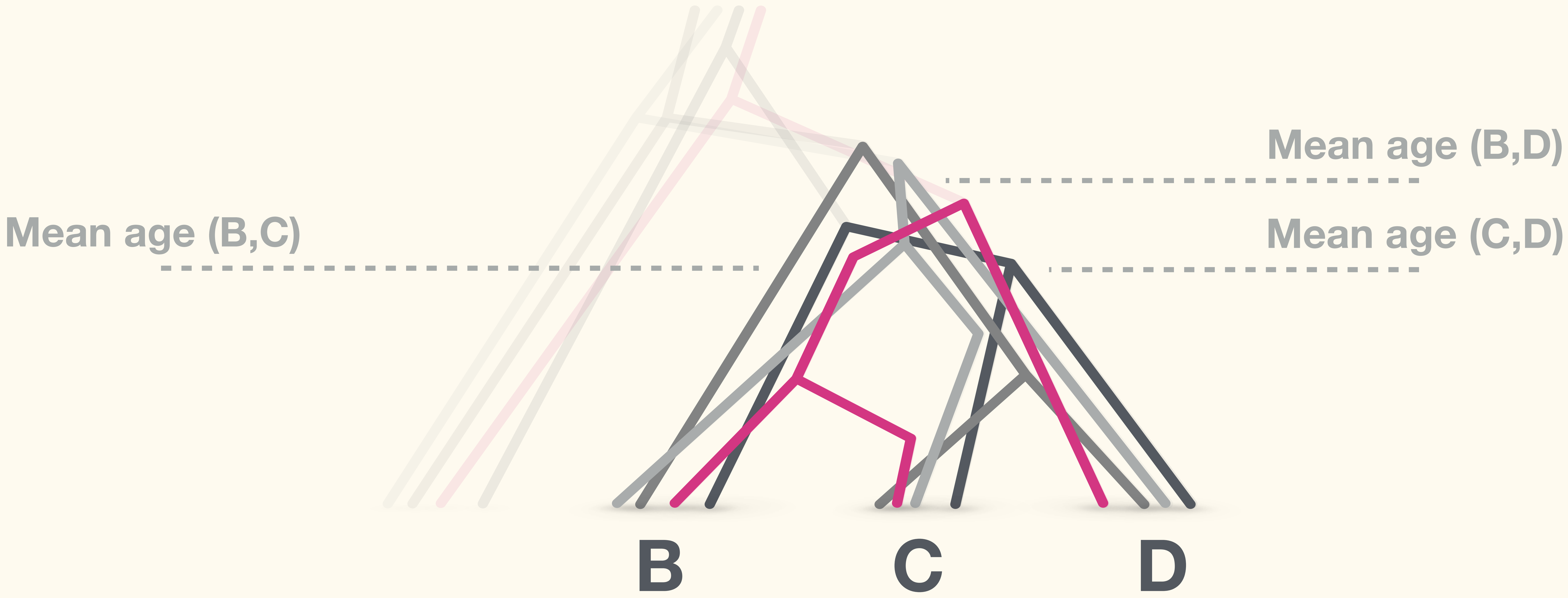
Introgression tests

Divergence times



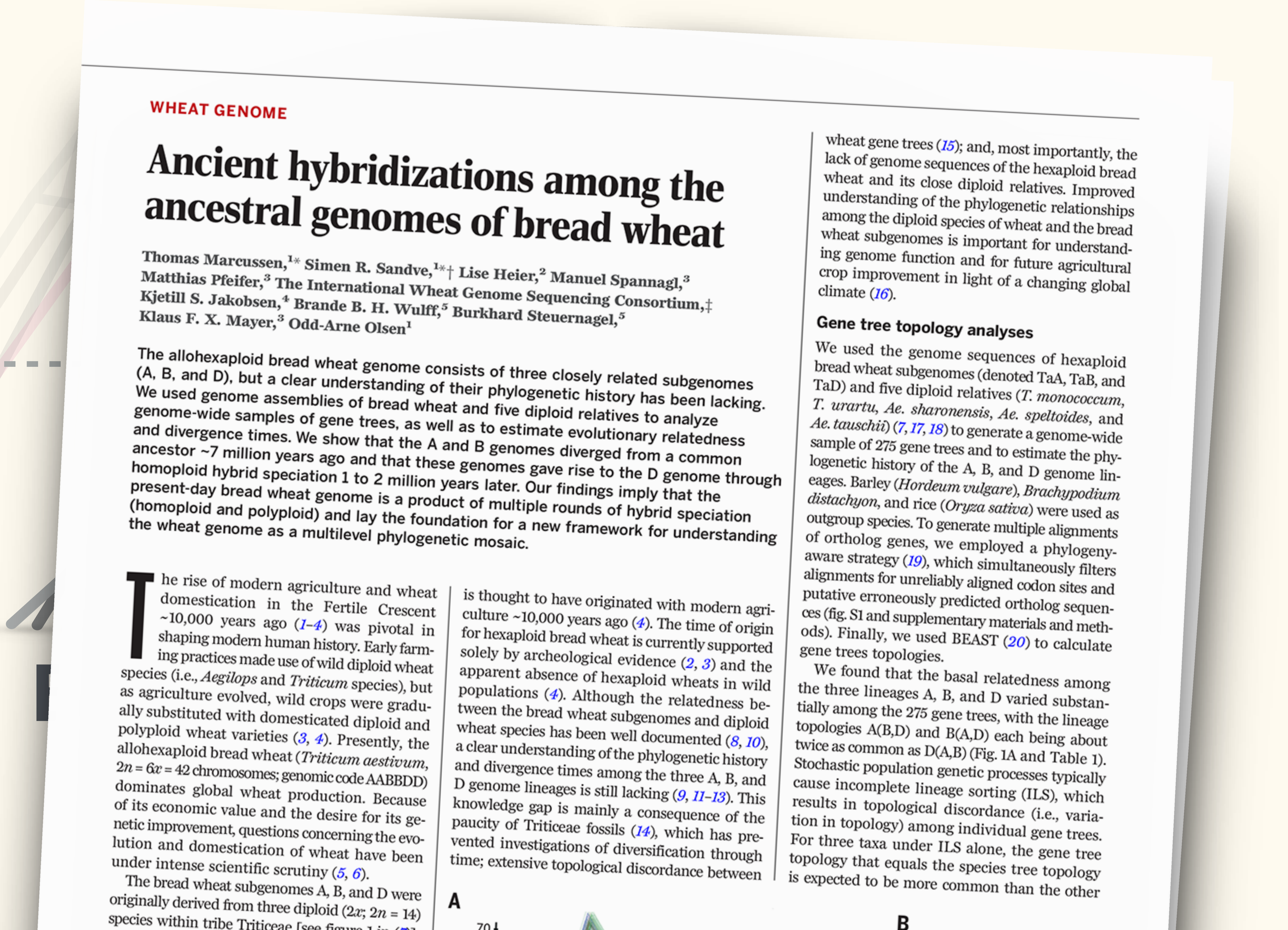
Introgression tests

Divergence times



Introgression tests

Divergence times



WHEAT GENOME

Ancient hybridizations among the ancestral genomes of bread wheat

Thomas Marcussen,[1*] Simen R. Sandve,[1*†] Lise Heier,[2] Manuel Spannagl,[3] Matthias Pfeifer,[3] The International Wheat Genome Sequencing Consortium,‡ Kjetill S. Jakobsen,[4] Brande B. H. Wulff,[5] Burkhard Steuernagel,[5] Klaus F. X. Mayer,[3] Odd-Arne Olsen[1]

The allohexaploid bread wheat genome consists of three closely related subgenomes (A, B, and D), but a clear understanding of their phylogenetic history has been lacking. We used genome assemblies of bread wheat and five diploid relatives to analyze genome-wide samples of gene trees, as well as to estimate evolutionary relatedness and divergence times. We show that the A and B genomes diverged from a common ancestor ~7 million years ago and that these genomes gave rise to the D genome through homoploid hybrid speciation 1 to 2 million years later. Our findings imply that the present-day bread wheat genome is a product of multiple rounds of hybrid speciation (homoploid and polyploid) and lay the foundation for a new framework for understanding the wheat genome as a multilevel phylogenetic mosaic.

The rise of modern agriculture and wheat domestication in the Fertile Crescent ~10,000 years ago (1–4) was pivotal in shaping modern human history. Early farming practices made use of wild diploid wheat species (i.e., *Aegilops* and *Triticum* species), but as agriculture evolved, wild crops were gradually substituted with domesticated diploid and polyploid wheat varieties (3, 4). Presently, the allohexaploid bread wheat (*Triticum aestivum*, $2n = 6x = 42$ chromosomes; genomic code AABBDD) dominates global wheat production. Because of its economic value and the desire for its genetic improvement, questions concerning the evolution and domestication of wheat have been under intense scientific scrutiny (5, 6).

The bread wheat subgenomes A, B, and D were originally derived from three diploid ($2x$; $2n = 14$) species within tribe Triticeae [see figure 1 in

is thought to have originated with modern agriculture ~10,000 years ago (4). The time of origin for hexaploid bread wheat is currently supported solely by archeological evidence (2, 3) and the apparent absence of hexaploid wheats in wild populations (4). Although the relatedness between the bread wheat subgenomes and diploid wheat species has been well documented (8, 10), a clear understanding of the phylogenetic history and divergence times among the three A, B, and D genome lineages is still lacking (9, 11–13). This knowledge gap is mainly a consequence of the paucity of Triticeae fossils (14), which has prevented investigations of diversification through time; extensive topological discordance between wheat gene trees (15); and, most importantly, the lack of genome sequences of the hexaploid bread wheat and its close diploid relatives. Improved understanding of the phylogenetic relationships among the diploid species of wheat and the bread wheat subgenomes is important for understanding genome function and for future agricultural crop improvement in light of a changing global climate (16).

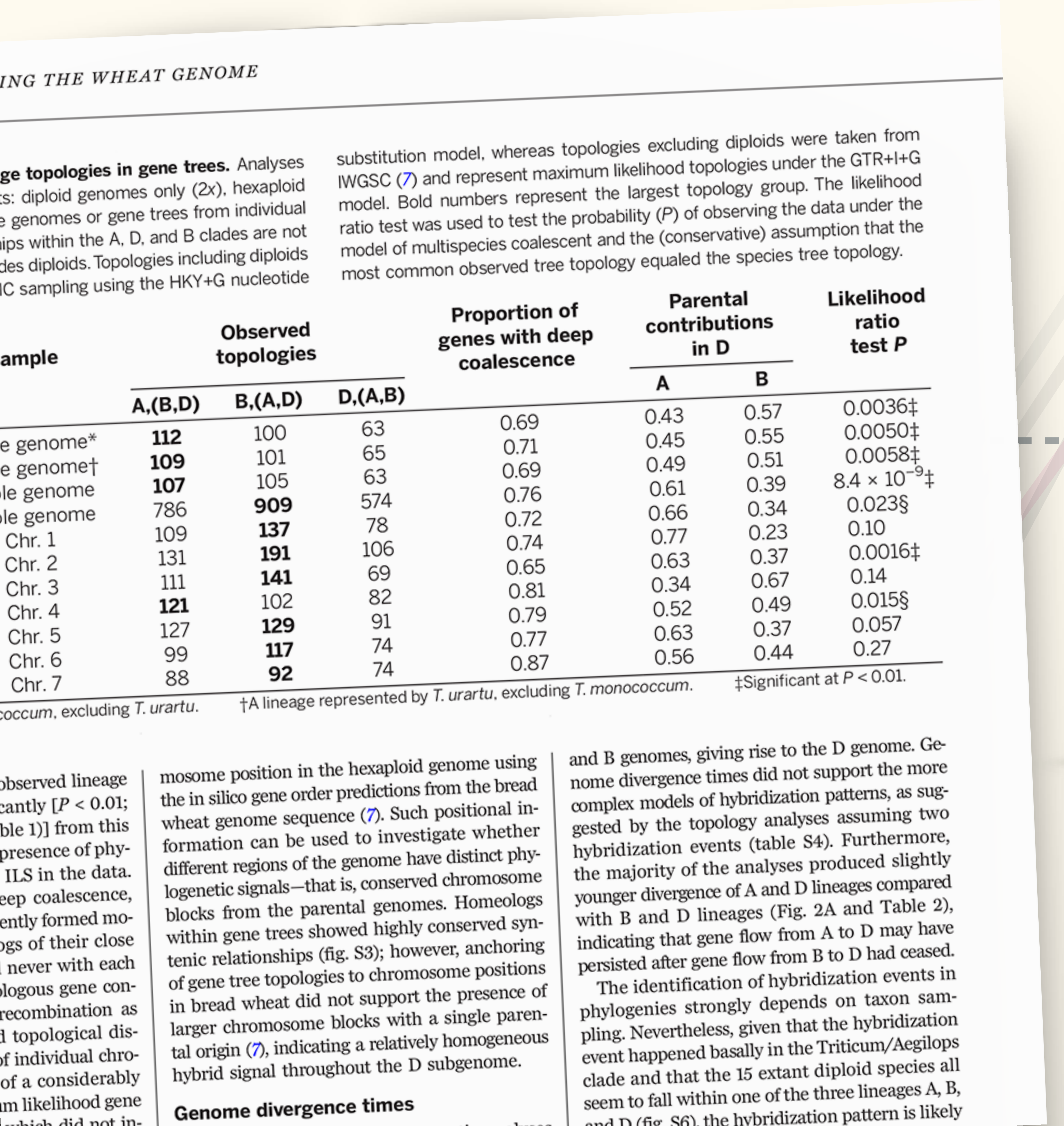
Gene tree topology analyses

We used the genome sequences of hexaploid bread wheat subgenomes (denoted TaA, TaB, and TaD) and five diploid relatives (*T. monococcum*, *T. urartu*, *Ae. sharonensis*, *Ae. speltoides*, and *Ae. tauschii*) (7, 17, 18) to generate a genome-wide sample of 275 gene trees and to estimate the phylogenetic history of the A, B, and D genome lineages. Barley (*Hordeum vulgare*), *Brachypodium distachyon*, and rice (*Oryza sativa*) were used as outgroup species. To generate multiple alignments of ortholog genes, we employed a phylogeny-aware strategy (19), which simultaneously filters alignments for unreliably aligned codon sites and putative erroneously predicted ortholog sequences (fig. S1 and supplementary materials and methods). Finally, we used BEAST (20) to calculate gene trees topologies.

We found that the basal relatedness among the three lineages A, B, and D varied substantially among the 275 gene trees, with the lineage topologies A(B,D) and B(A,D) each being about twice as common as D(A,B) (Fig. 1A and Table 1). Stochastic population genetic processes typically cause incomplete lineage sorting (ILS), which results in topological discordance (i.e., variation in topology) among individual gene trees. For three taxa under ILS alone, the gene tree topology that equals the species tree topology is expected to be more common than the other

Introgression tests

Divergence times



ING THE WHEAT GENOME

ge topologies in gene trees. Analyses ts: diploid genomes only (2x), hexaploid e genomes or gene trees from individual ips within the A, D, and B clades are not des diploids. Topologies including diploids IC sampling using the HKY+G nucleotide substitution model, whereas topologies excluding diploids were taken from IWGSC (7) and represent maximum likelihood topologies under the GTR+I+G model. Bold numbers represent the largest topology group. The likelihood ratio test was used to test the probability (*P*) of observing the data under the model of multispecies coalescent and the (conservative) assumption that the most common observed tree topology equaled the species tree topology.

ample	Observed topologies			Proportion of genes with deep coalescence	Parental contributions in D		Likelihood ratio test *P*
	A,(B,D)	B,(A,D)	D,(A,B)		A	B	
e genome*	**112**	100	63	0.69	0.43	0.57	0.0036‡
e genome†	**109**	101	65	0.71	0.45	0.55	0.0050‡
le genome	**107**	105	63	0.69	0.49	0.51	0.0058‡
le genome	786	**909**	574	0.76	0.61	0.39	8.4×10^{-9}‡
Chr. 1	109	**137**	78	0.72	0.66	0.34	0.023§
Chr. 2	131	**191**	106	0.74	0.77	0.23	0.10
Chr. 3	111	**141**	69	0.65	0.63	0.37	0.0016‡
Chr. 4	**121**	102	82	0.81	0.34	0.67	0.14
Chr. 5	127	**129**	91	0.79	0.52	0.49	0.015§
Chr. 6	99	**117**	74	0.77	0.63	0.37	0.057
Chr. 7	88	**92**	74	0.87	0.56	0.44	0.27

coccum, excluding *T. urartu*. †A lineage represented by *T. urartu*, excluding *T. monococcum*. ‡Significant at *P* < 0.01.

observed lineage cantly [*P* < 0.01; ble 1)] from this presence of phy- ILS in the data. eep coalescence, ently formed mo- gs of their close never with each logous gene con- recombination as d topological dis- f individual chro- of a considerably m likelihood gene

mosome position in the hexaploid genome using the in silico gene order predictions from the bread wheat genome sequence (7). Such positional information can be used to investigate whether different regions of the genome have distinct phylogenetic signals—that is, conserved chromosome blocks from the parental genomes. Homeologs within gene trees showed highly conserved syntenic relationships (fig. S3); however, anchoring of gene tree topologies to chromosome positions in bread wheat did not support the presence of larger chromosome blocks with a single parental origin (7), indicating a relatively homogeneous hybrid signal throughout the D subgenome.

Genome divergence times

and B genomes, giving rise to the D genome. Genome divergence times did not support the more complex models of hybridization patterns, as suggested by the topology analyses assuming two hybridization events (table S4). Furthermore, the majority of the analyses produced slightly younger divergence of A and D lineages compared with B and D lineages (Fig. 2A and Table 2), indicating that gene flow from A to D may have persisted after gene flow from B to D had ceased.

The identification of hybridization events in phylogenies strongly depends on taxon sampling. Nevertheless, given that the hybridization event happened basally in the Triticum/Aegilops clade and that the 15 extant diploid species all seem to fall within one of the three lineages A, B, and D (fig. S6), the hybridization pattern is likely

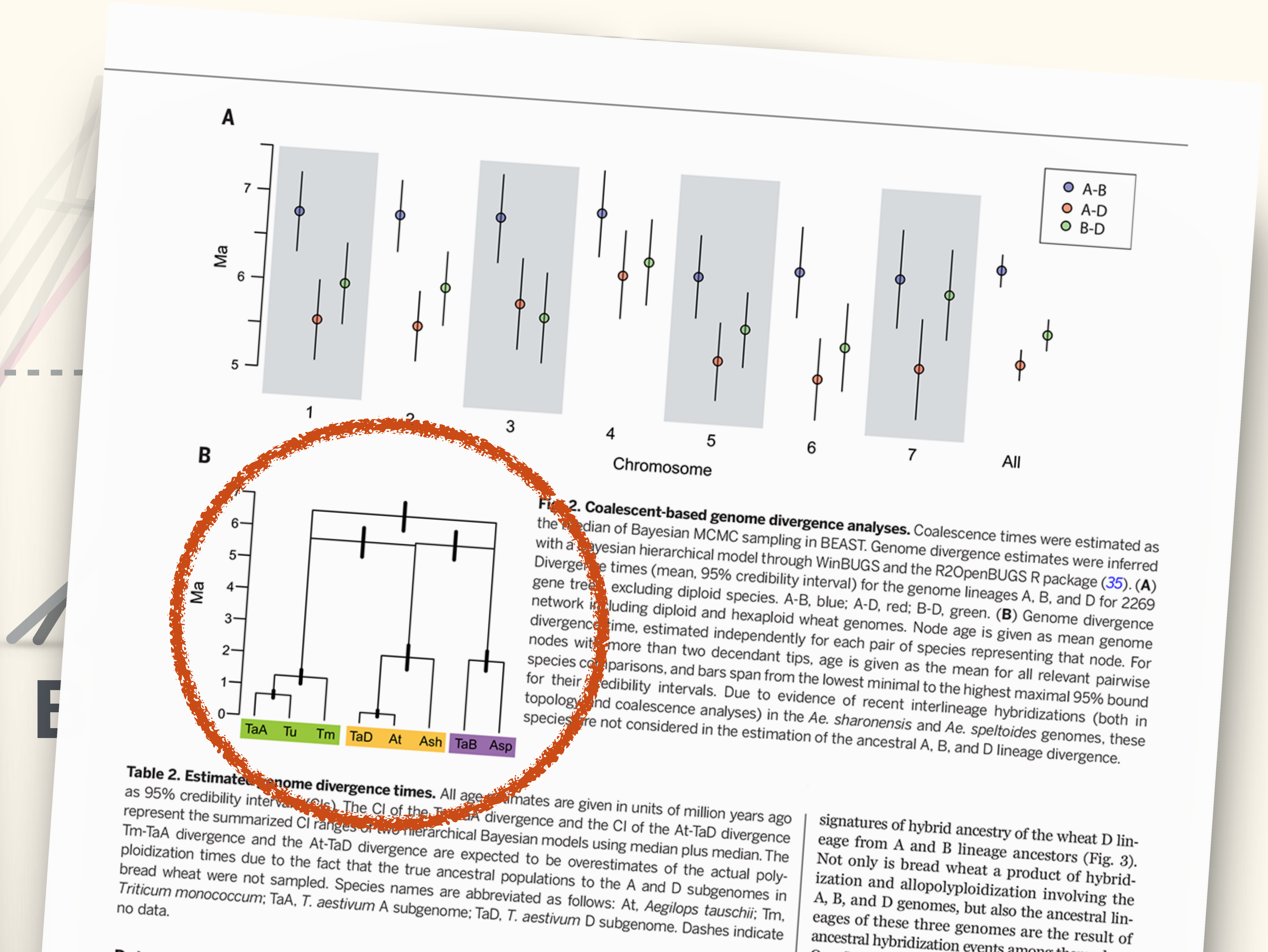


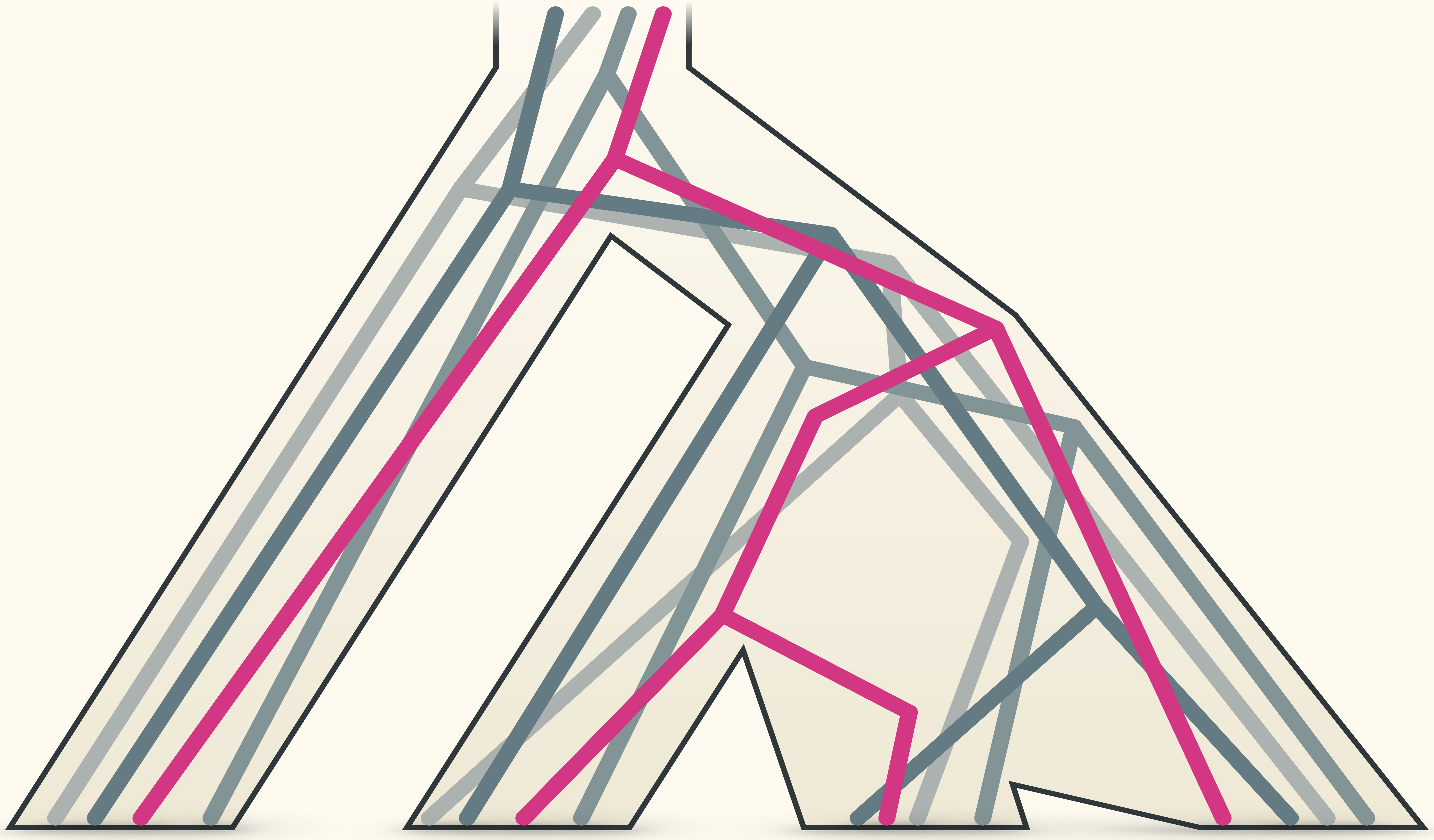
Fig. 2. Coalescent-based genome divergence analyses. Coalescence times were estimated as the median of Bayesian MCMC sampling in BEAST. Genome divergence estimates were inferred with a Bayesian hierarchical model through WinBUGS and the R2OpenBUGS R package (35). (A) Divergence times (mean, 95% credibility interval) for the genome lineages A, B, and D for 2269 gene trees excluding diploid species. A-B, blue; A-D, red; B-D, green. (B) Genome divergence network including diploid and hexaploid wheat genomes. Node age is given as mean genome divergence time, estimated independently for each pair of species representing that node. For nodes with more than two decendant tips, age is given as the mean for all relevant pairwise species comparisons, and bars span from the lowest minimal to the highest maximal 95% bound for their credibility intervals. Due to evidence of recent interlineage hybridizations (both in topology and coalescence analyses) in the *Ae. sharonensis* and *Ae. speltoides* genomes, these species were not considered in the estimation of the ancestral A, B, and D lineage divergence.

Table 2. Estimated genome divergence times. All age estimates are given in units of million years ago as 95% credibility intervals (CIs). The CI of the Tm-TaA divergence and the CI of the At-TaD divergence represent the summarized CI ranges of two hierarchical Bayesian models using median plus median. The Tm-TaA divergence and the At-TaD divergence are expected to be overestimates of the actual polyploidization times due to the fact that the true ancestral populations to the A and D subgenomes in bread wheat were not sampled. Species names are abbreviated as follows: At, *Aegilops tauschii*; Tm, *Triticum monococcum*; TaA, *T. aestivum* A subgenome; TaD, *T. aestivum* D subgenome. Dashes indicate no data.

signatures of hybrid ancestry of the wheat D lineage from A and B lineage ancestors (Fig. 3). Not only is bread wheat a product of hybridization and allopolyploidization involving the A, B, and D genomes, but also the ancestral lineages of these three genomes are the result of ancestral hybridization events among

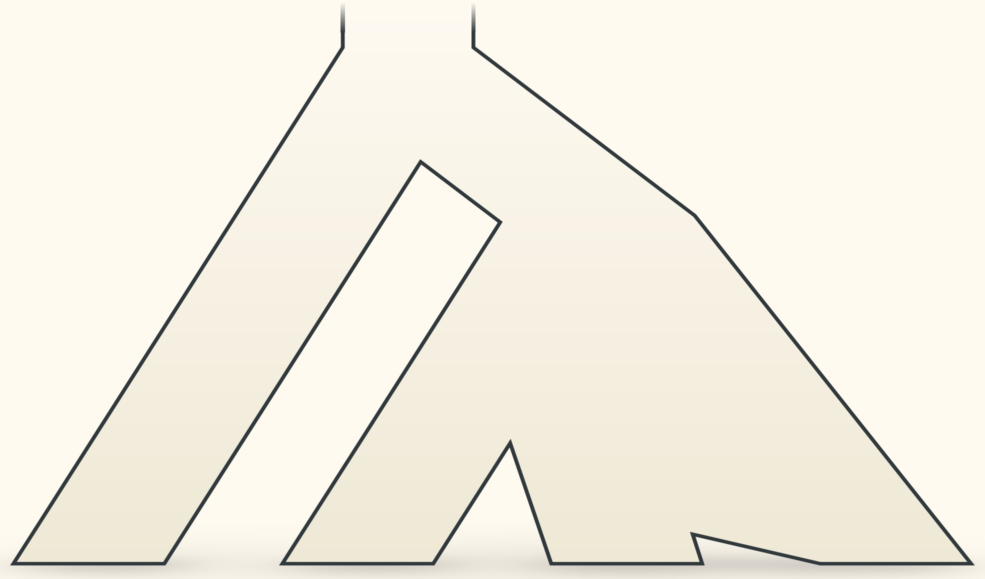
Introgression tests

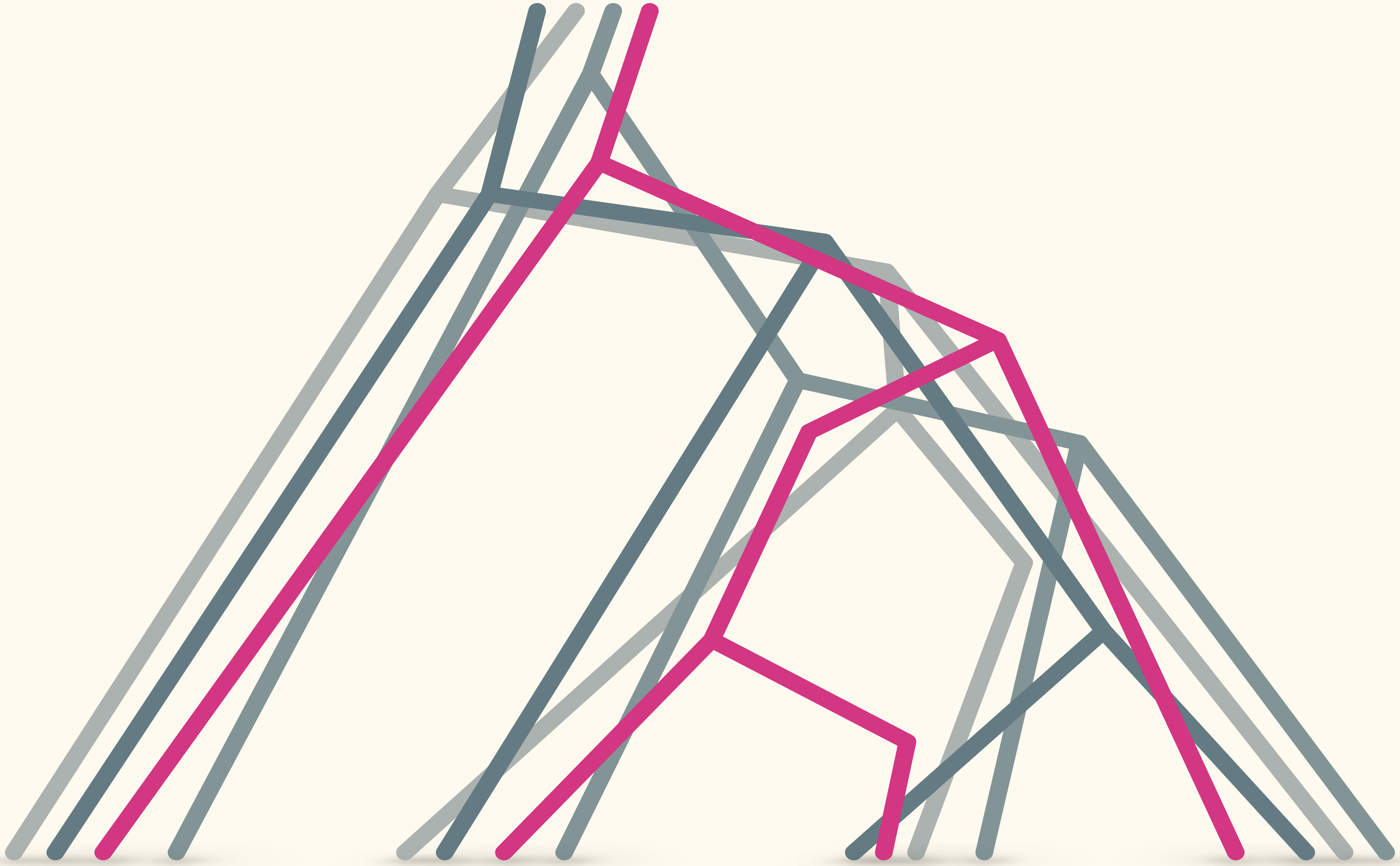
Model-based inference



Introgression tests

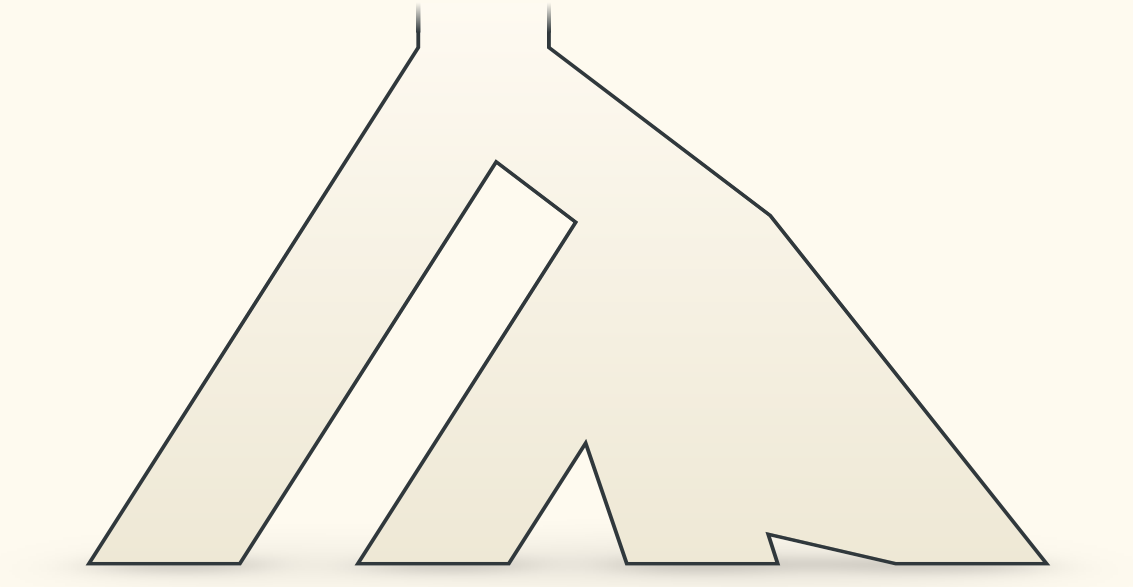
Model-based inference

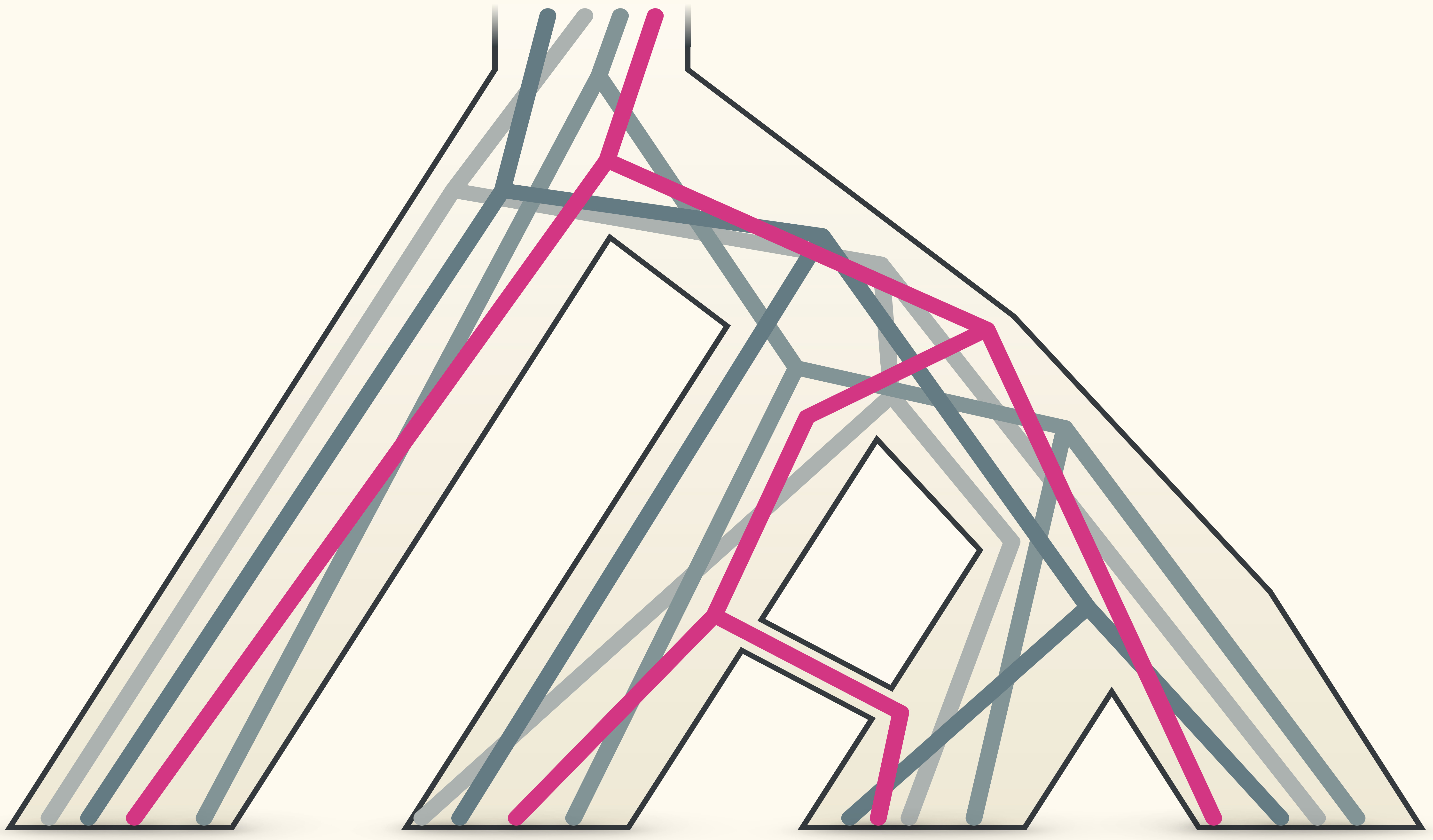




Introgression tests

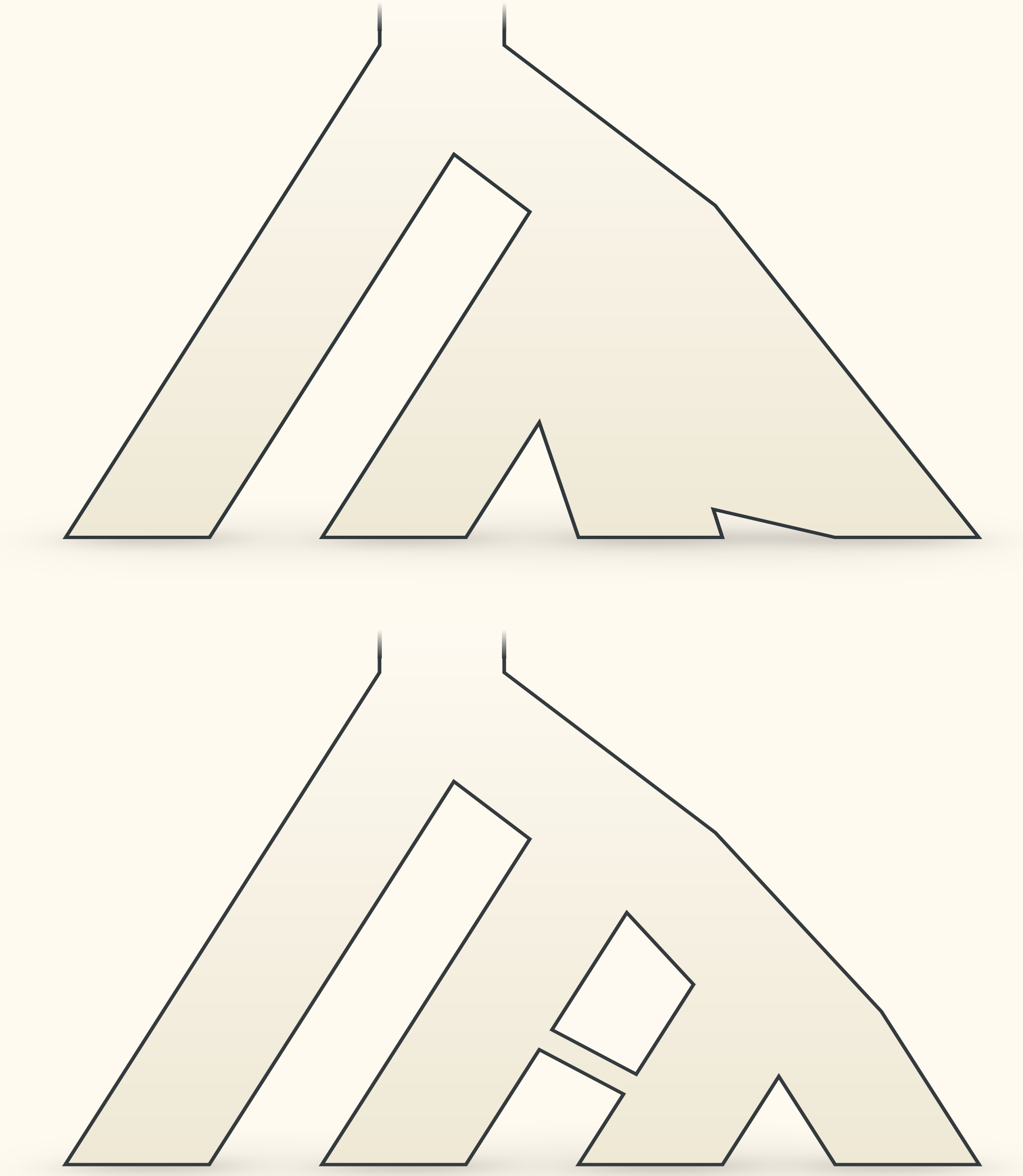
Model-based inference

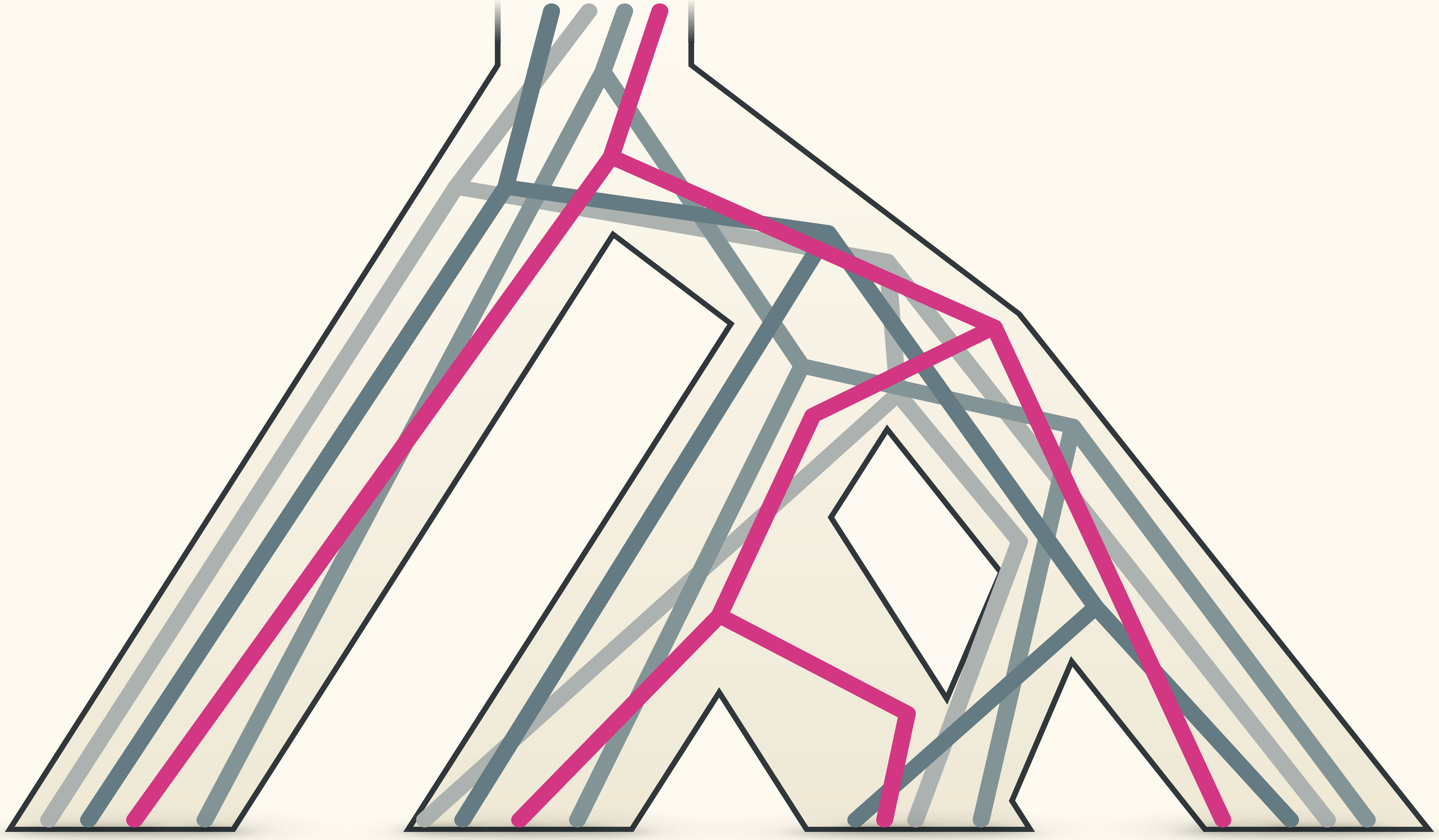




Introgression tests

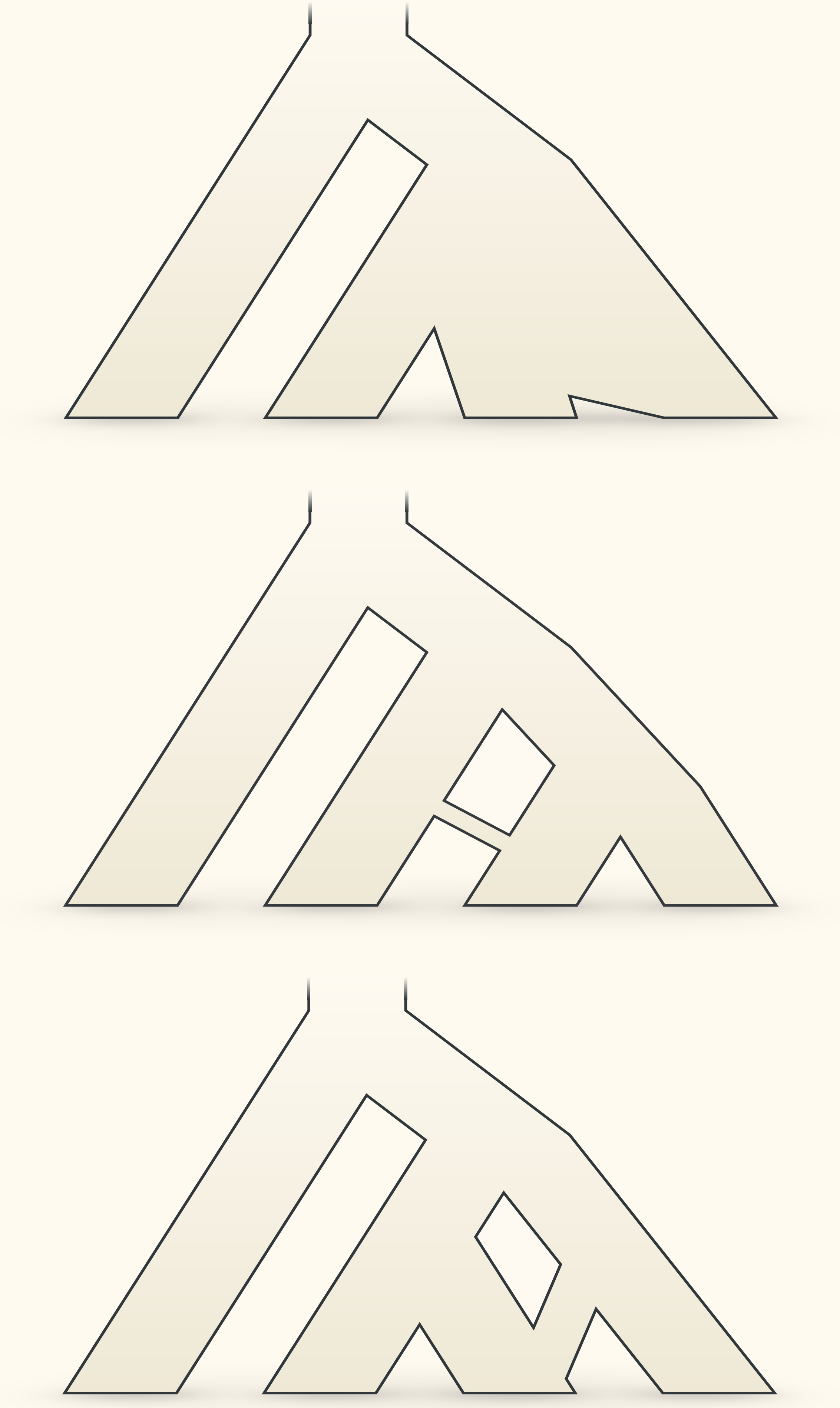
Model-based inference

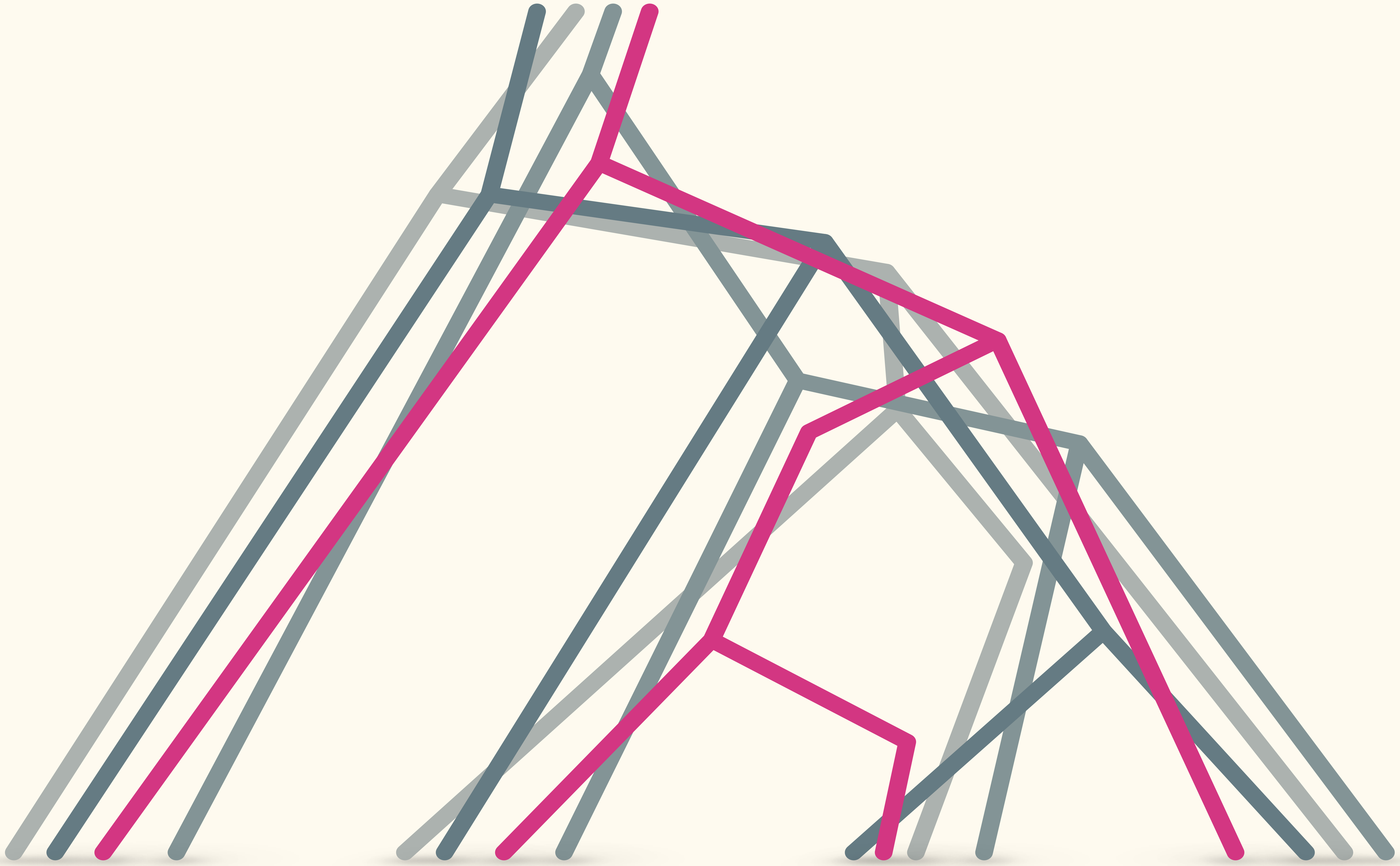




Introgression tests

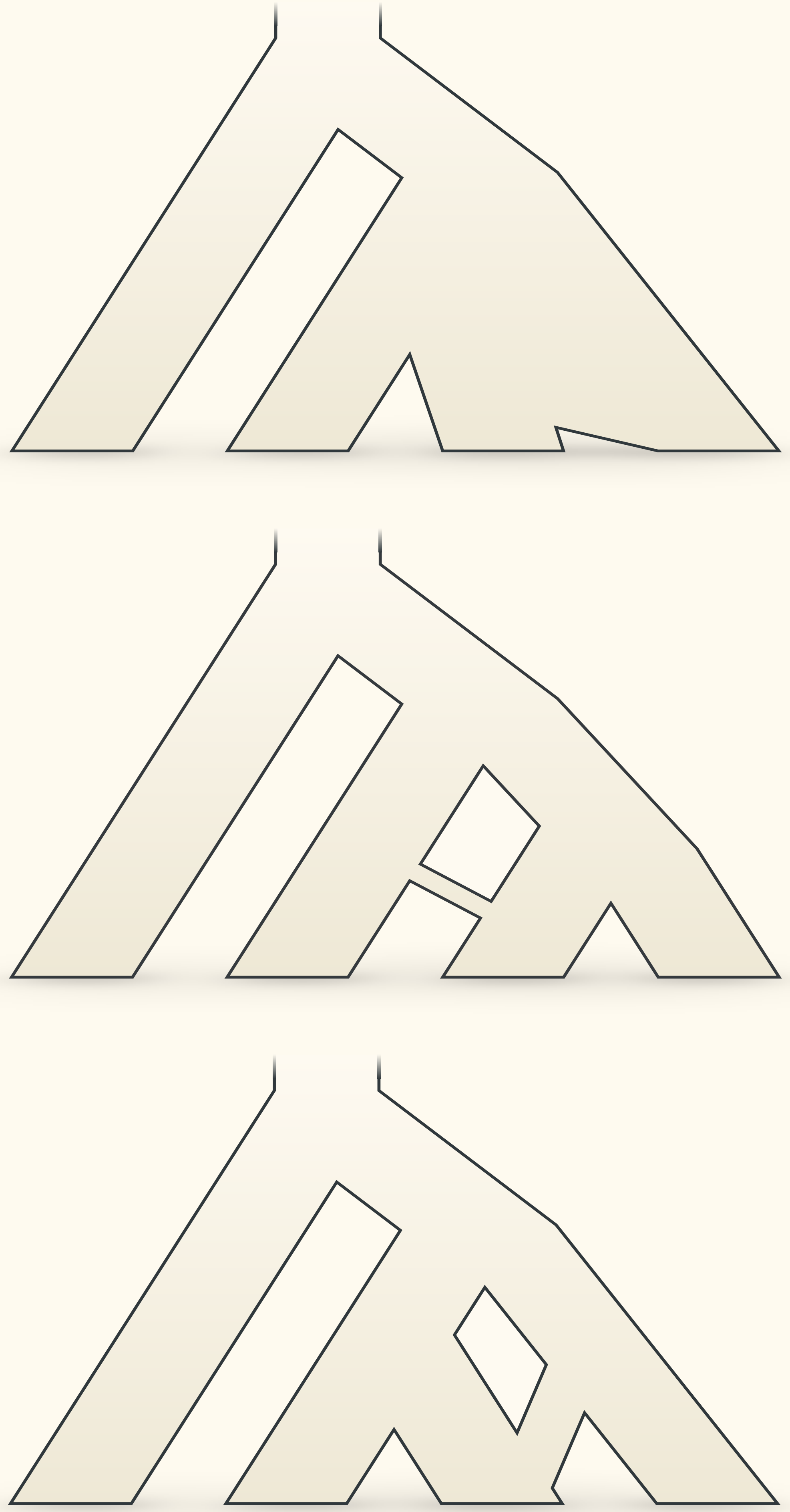
Model-based inference

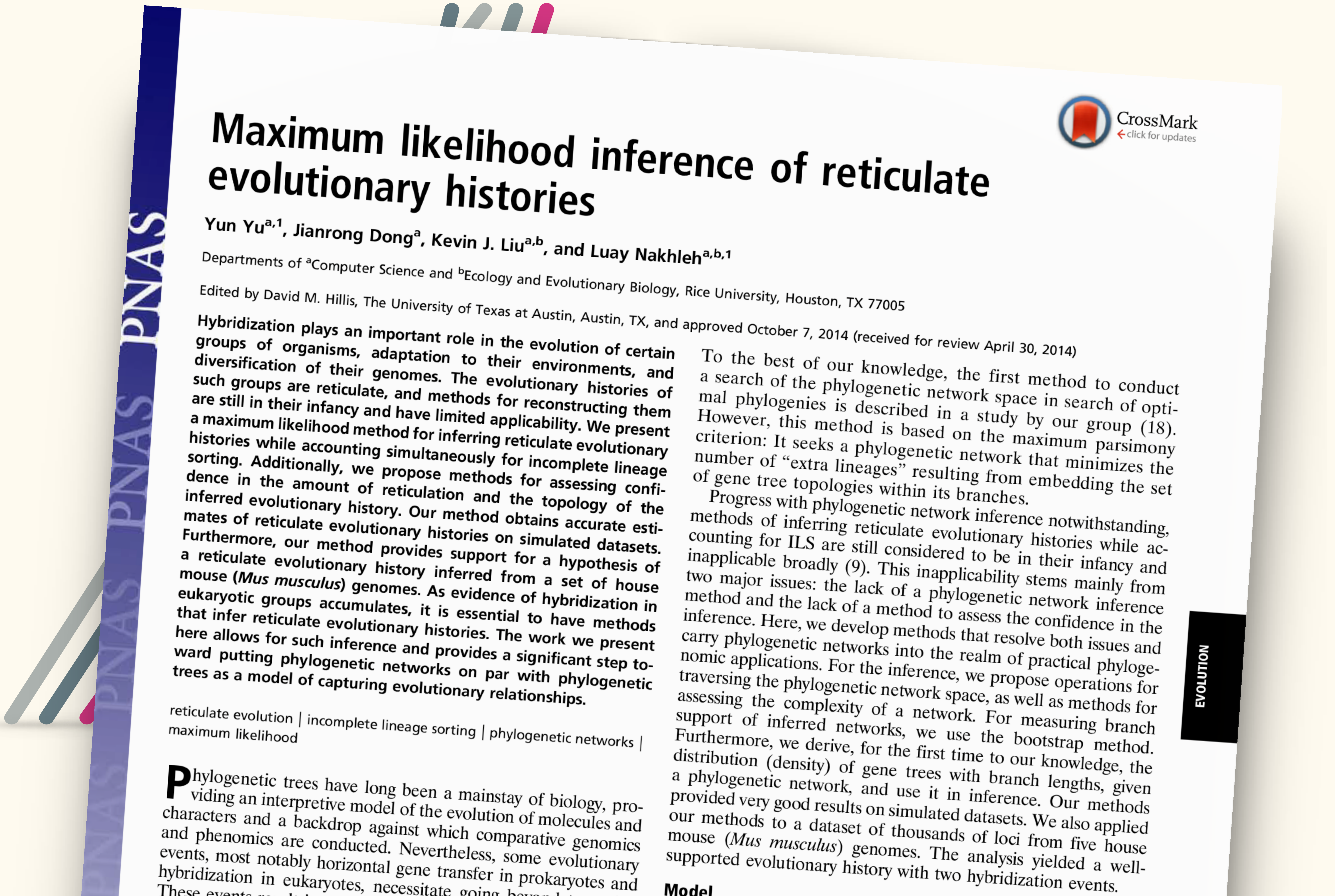




Introgression tests

Model-based inference





Maximum likelihood inference of reticulate evolutionary histories

Yun Yu[a,1], Jianrong Dong[a], Kevin J. Liu[a,b], and Luay Nakhleh[a,b,1]

Departments of [a]Computer Science and [b]Ecology and Evolutionary Biology, Rice University, Houston, TX 77005

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved October 7, 2014 (received for review April 30, 2014)

Hybridization plays an important role in the evolution of certain groups of organisms, adaptation to their environments, and diversification of their genomes. The evolutionary histories of such groups are reticulate, and methods for reconstructing them are still in their infancy and have limited applicability. We present a maximum likelihood method for inferring reticulate evolutionary histories while accounting simultaneously for incomplete lineage sorting. Additionally, we propose methods for assessing confidence in the amount of reticulation and the topology of the inferred evolutionary history. Our method obtains accurate estimates of reticulate evolutionary histories on simulated datasets. Furthermore, our method provides support for a hypothesis of a reticulate evolutionary history inferred from a set of house mouse (*Mus musculus*) genomes. As evidence of hybridization in eukaryotic groups accumulates, it is essential to have methods that infer reticulate evolutionary histories. The work we present here allows for such inference and provides a significant step toward putting phylogenetic networks on par with phylogenetic trees as a model of capturing evolutionary relationships.

reticulate evolution | incomplete lineage sorting | phylogenetic networks | maximum likelihood

Phylogenetic trees have long been a mainstay of biology, providing an interpretive model of the evolution of molecules and characters and a backdrop against which comparative genomics and phenomics are conducted. Nevertheless, some evolutionary events, most notably horizontal gene transfer in prokaryotes and hybridization in eukaryotes, necessitate going

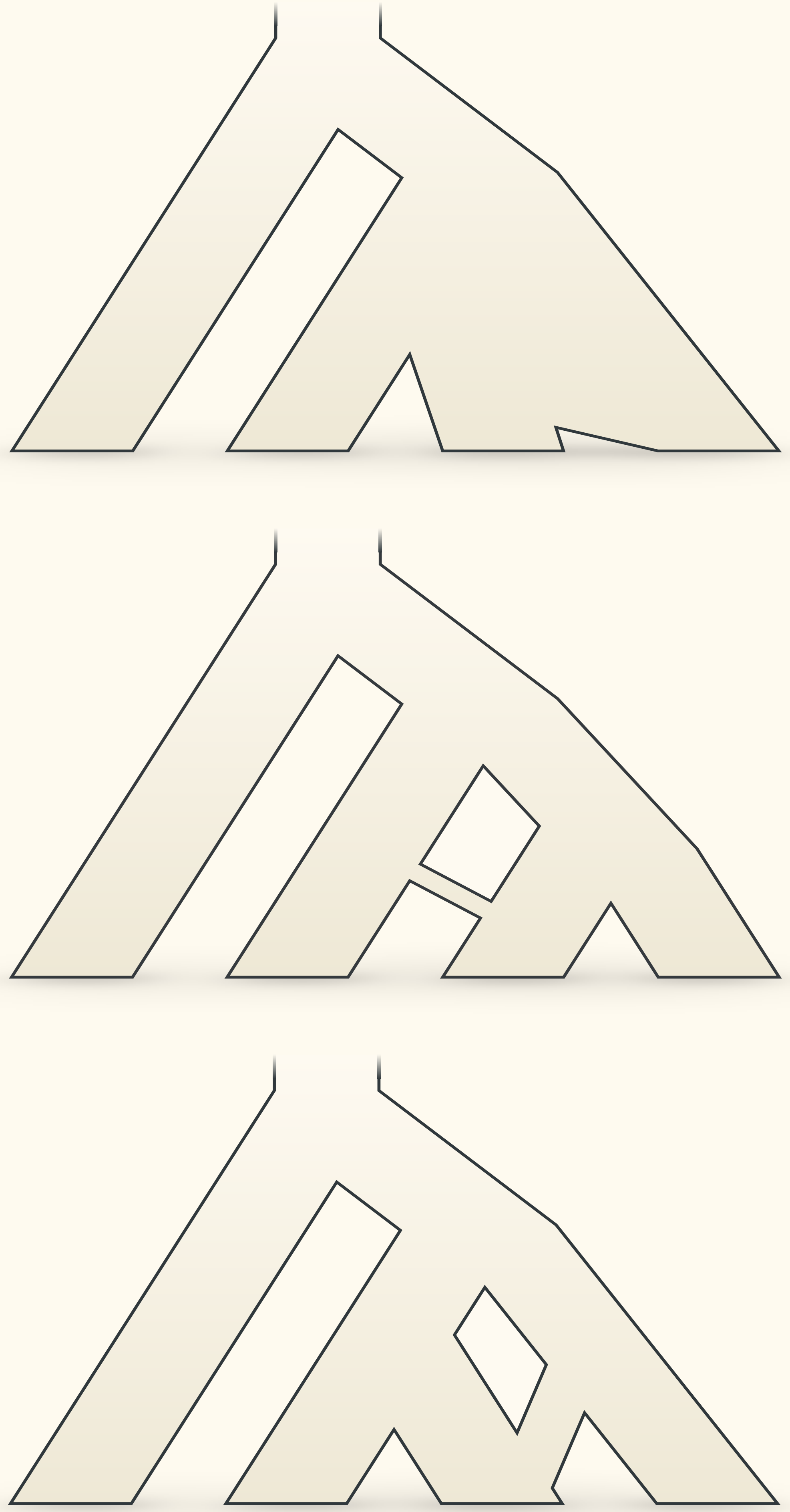
To the best of our knowledge, the first method to conduct a search of the phylogenetic network space in search of optimal phylogenies is described in a study by our group (18). However, this method is based on the maximum parsimony criterion: It seeks a phylogenetic network that minimizes the number of "extra lineages" resulting from embedding the set of gene tree topologies within its branches.

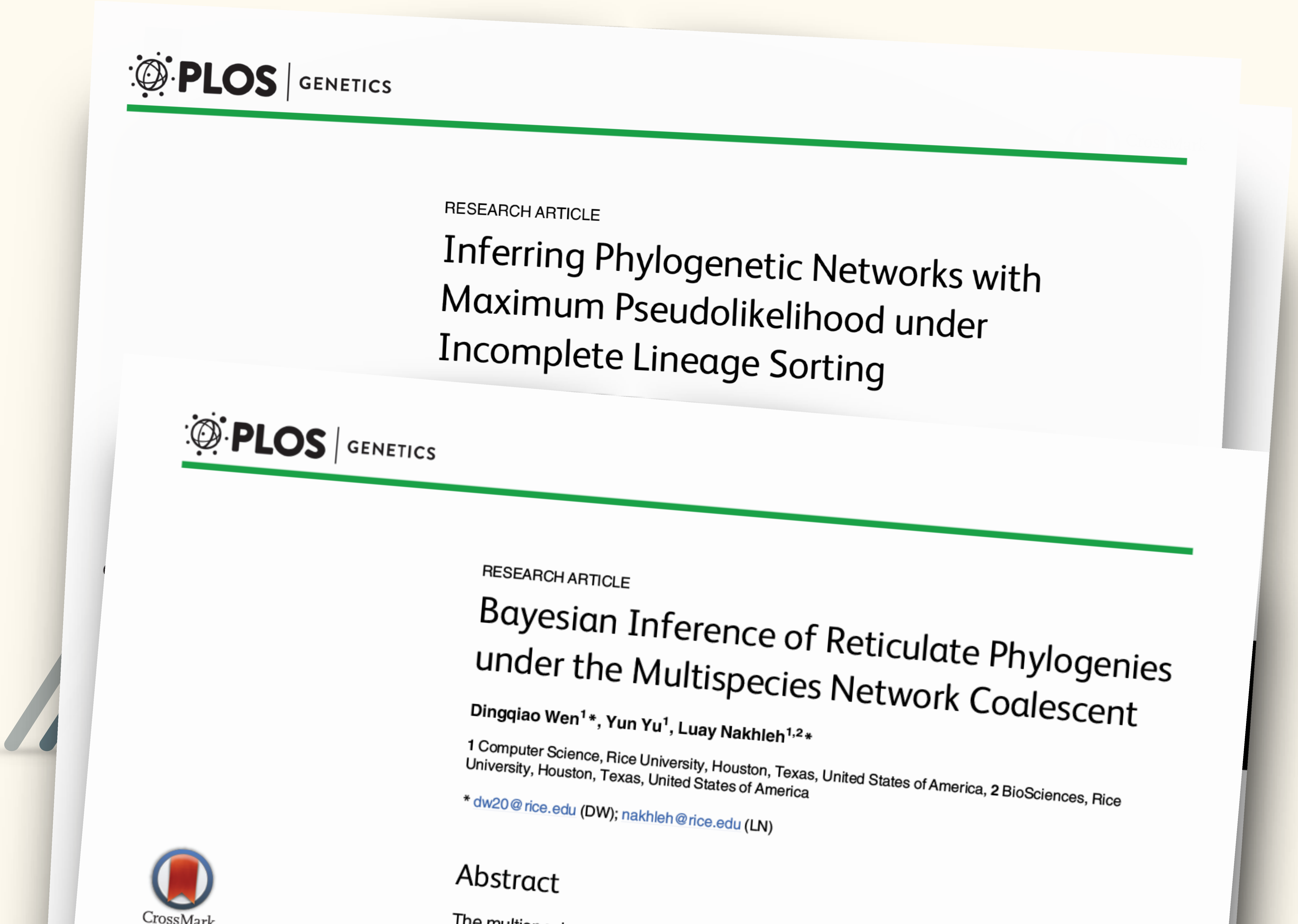
Progress with phylogenetic network inference notwithstanding, methods of inferring reticulate evolutionary histories while accounting for ILS are still considered to be in their infancy and inapplicable broadly (9). This inapplicability stems mainly from two major issues: the lack of a phylogenetic network inference method and the lack of a method to assess the confidence in the inference. Here, we develop methods that resolve both issues and carry phylogenetic networks into the realm of practical phylogenomic applications. For the inference, we propose operations for traversing the phylogenetic network space, as well as methods for assessing the complexity of a network. For measuring branch support of inferred networks, we use the bootstrap method. Furthermore, we derive, for the first time to our knowledge, the distribution (density) of gene trees with branch lengths, given a phylogenetic network, and use it in inference. Our methods provided very good results on simulated datasets. We also applied our methods to a dataset of thousands of loci from five house mouse (*Mus musculus*) genomes. The analysis yielded a well-supported evolutionary history with two hybridization events.

Model

Introgression tests

Model-based inference





PLOS | GENETICS

RESEARCH ARTICLE

Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting

PLOS | GENETICS

RESEARCH ARTICLE

Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent

Dingqiao Wen[1]*, Yun Yu[1], Luay Nakhleh[1,2]*

[1] Computer Science, Rice University, Houston, Texas, United States of America, [2] BioSciences, Rice University, Houston, Texas, United States of America

* dw20@rice.edu (DW); nakhleh@rice.edu (LN)

Abstract

Thanks