Using conventional *F*-statistics to study unconventional sex-chromosome differentiation

Nicolas Rodrigues and Christophe Dufresnes

Department of Ecology & Evolution, University of Lausanne, Lausanne, Switzerland

ABSTRACT

Species with undifferentiated sex chromosomes emerge as key organisms to understand the astonishing diversity of sex-determination systems. Whereas new genomic methods are widening opportunities to study these systems, the difficulty to separately characterize their X and Y homologous chromosomes poses limitations. Here we demonstrate that two simple *F*-statistics calculated from sex-linked genotypes, namely the genetic distance (F_{st}) between sexes and the inbreeding coefficient (F_{is}) in the heterogametic sex, can be used as reliable proxies to compare sex-chromosome differentiation between populations. We correlated these metrics using published microsatellite data from two frog species (*Hyla arborea* and *Rana temporaria*), and show that they intimately relate to the overall amount of X–Y differentiation in populations. However, the fits for individual loci appear highly variable, suggesting that a dense genetic coverage will be needed for inferring fine-scale patterns of differentiation along sex-chromosomes. The applications of these *F*-statistics, which implies little sampling requirement, significantly facilitate population analyses of sex-chromosomes.

INTRODUCTION

In sharp contrast with the classical sex-determining systems of mammals and birds, the study of sex-chromosome evolution in other vertebrate lineages has revealed a myriad of alternative evolutionary trajectories (*Beukeboom & Perrin, 2014*). Species with homomorphic gametologs are providing instrumental insights into the mechanisms paving these unconventional pathways, like the rates of sex-chromosome transitions (e.g., *Dufresnes et al., 2015*), the dynamics of X–Y recombination (e.g., *Stöck et al., 2013*; *Dufresnes et al., 2014b*), the evolution of X–Y differentiation (e.g., *Yoshida et al., 2014*), as well as the interplay between genetic and non-genetic sex-determination (e.g., *Rodrigues et al., 2015*; *Perrin, 2016*). Often neglected due to the lack of genomic resources, these promising non-model organisms can now be widely exploited for sex-chromosome research with low-cost population genomic techniques (*Brelsford, Dufresnes & Perrin, 2016a*; *Brelsford et al., in press*). However, given the rapid evolution of the forces at work, patterns of variation at sex-linked markers can be complex and population-specific (*Rodrigues et al., 2014*; *Dufresnes et*

Corresponding authors
Nicolas Rodrigues,
Nicolas.Rodrigues@unil.ch
Christophe Dufresnes,
Christophe.Dufresnes@unil.ch

al., 2014a; *Dufresnes et al., 2014b*), prompting for multilevel analyses in order to get comprehensive inferences.

A key variable to such analyses is the amount of differentiation between sex chromosomes. This feature, central to the evolutionary history of sex chromosomes, is highly informative regarding their contribution to sex-determination, how they differentiate and which genomic regions are affected. For instance, mapping peaks of X–Y divergence can point to sex-determining regions (e.g., *Brelsford, Dufresnes & Perrin, 2016b*); in a similar fashion, it can be used to screen for sex-antagonistic genes and thus test their hypothetical role in triggering the suppression of X–Y recombination (*Kirkpatrick & Guerrero, 2014*), a critical and criticized assumption in the sex-chromosome literature (*Beukeboom & Perrin, 2014*; *Wright et al., 2016*).

Measuring sex-chromosome differentiation in species with "undifferentiated" sex chromosomes is by definition challenging. Unlike in mammals and birds, these sex chromosomes are largely homologous. Thus, estimating genetic divergence between the X and Y copies of homologous loci requires their separate genotyping (by cloning methods), or to phase X and Y haplotypes in males from patterns of linkage disequilibrium. Both of these approaches have severe limitations for population genetics and phylogeographic analyses. Cloning is only adequate for genotyping few genes in few individuals. Phasing diploid genotypes requires tremendous sampling and genotyping efforts, including large adult (males and females) and family samples (crosses) in populations. Moreover, given that it relies on linkage disequilibrium, the latter is easier and thus biased towards populations where XY recombination is low or null (and XY differentiation is high). Already challenging with small datasets like microsatellite genotypes, haplotype reconstruction becomes a struggle with high-throughput genomic data.

An indirect ad hoc alternative is to compute allele frequency indices on sexed samples, like F-statistics. Genetic distance between males and females from a panmictic population should be proportional to the amount of X–Y differentiation. Because males share half of their sex-linked alleles with females (the X copies), pairwise F_{st} between sexes ($\male-\female F_{st}$) is thus expected to span from 0.0 (null X–Y differentiation) to 0.5 (complete X–Y differentiation). Even simpler, X–Y differentiation can theoretically be quantified through the excesses of heterozygotes at sex-linked loci in the heterogametic sex, i.e., XY males, thus without the systematic need for female samples. Heterozygote excess is commonly depicted by negative F_{is} values. Hence, male F_{is} ($\male F_{is}$) at sex-linked loci should span from 0.0 (no X–Y differentiation) to -1.0 (complete X–Y differentiation) in populations at Hardy–Weinberg Equilibrium (HWE). The rationales of these ad hoc approaches appear straightforward and have been used in few previous studies (e.g., *Shikano et al., 2011*; *Natri, Shikano & Merilä, 2013*; *Dufresnes et al., 2014b*; *Rodrigues et al., 2014*). However, these F-statistics may also be influenced by other processes such as sex-specific dispersal, departure from HWE due to demographic processes, as well as drift shaping marker-specific signals, all of which may temper their reliability to estimate sex-chromosome differentiation. Thus, encouraging their application first necessitates proper assessment in comprehensive population genetic frameworks.

Here we demonstrate the informativeness of $\male - \female F_{st}$ and $\male F_{is}$ at sex-linked markers to reliably compare sex-chromosome differentiation between natural populations. We extracted and correlated these statistics from published microsatellite datasets of two famous study systems in the field of sex determination: the male-heterogametic frogs *Hyla arborea* and *Rana temporaria*, for which data from multiple populations are available for such comparison. The little requirements of these methods significantly enlarge opportunities for the study of homomorphic sex chromosomes in a wide array of non-model organisms.

METHODS

Hyla arborea data

This dataset includes sex-linked microsatellite genotypes across the entire range of the species in Europe, used to understand the evolution of X–Y differentiation and recombination in a phylogeographic framework (*Dufresnes et al., 2014b*; dryad doi: http://dx.doi.org/10.5061/dryad.45j84). To this end, using male and female adult samples (distinguished based on secondary sexual traits, i.e., the presence/absence of vocal sacs on the throat), combined with family data (parents + offspring), the authors could phase X and Y haplotypes for 11 microsatellite loci (details in *Dufresnes et al., 2014b*) across 28 populations of at least 5 males, and computed a metric of X–Y differentiation based on allele frequency overlap (described in *Dufresnes et al., 2014b*; page 3447). We extracted this data and computed $\male F_{is}$ for these populations using FSTAT (*Goudet, 1995*). We also calculated F_{st} between sexes ($\male - \female F_{st}$) for a subset of 14 of these populations, where at least five individuals of each sex were available (Table S1A). Sample size of less than five individuals were not considered in order to include only statistically robust estimates.

Moreover, in order to account for the baseline levels of inbreeding (see 'Results & Discussion'), we estimated the F_{is} of females at sex-linked loci ($\female F_{is}$). For the same purpose, we mined a second published dataset to compute F_{is} from autosomal microsatellite genotypes (autosomal F_{is}), which are available for 27 out of the 28 populations (*Dufresnes et al., 2013*; dryad doi: http://dx.doi.org/10.5061/dryad.2vk30; 30 loci). We then adjusted $\male F_{is}$ by computing the difference with either $\female F_{is}$ or autosomal F_{is}.

For each comparison, we fitted linear regression models in R (*R Core Team, 2016*).

Rana temporaria data

This dataset includes microsatellite genotypes (11–13 loci) of the sex-linkage group from six Swedish and four Swiss populations of at least five individuals of each sex (*Rodrigues et al., 2013*; dryad doi: http://dx.doi.org/10.5061/dryad.0mg7h; *Rodrigues et al., 2014*; dryad doi: http://dx.doi.org/10.5061/dryad.mb06v). This data was originally generated to investigate levels of sex-specific genetic differentiation at this linkage group to assess the relative contribution of genetic *vs.* non-genetic components of sex-determination in this species. As for *H. arborea*, we computed $\male F_{is}$, $\male - \female F_{st}$ as well as $\female F_{is}$ for each population (Table S1B), and fitted linear regression models. However, no measure of X–Y differentiation nor autosomal variation is available for these populations.
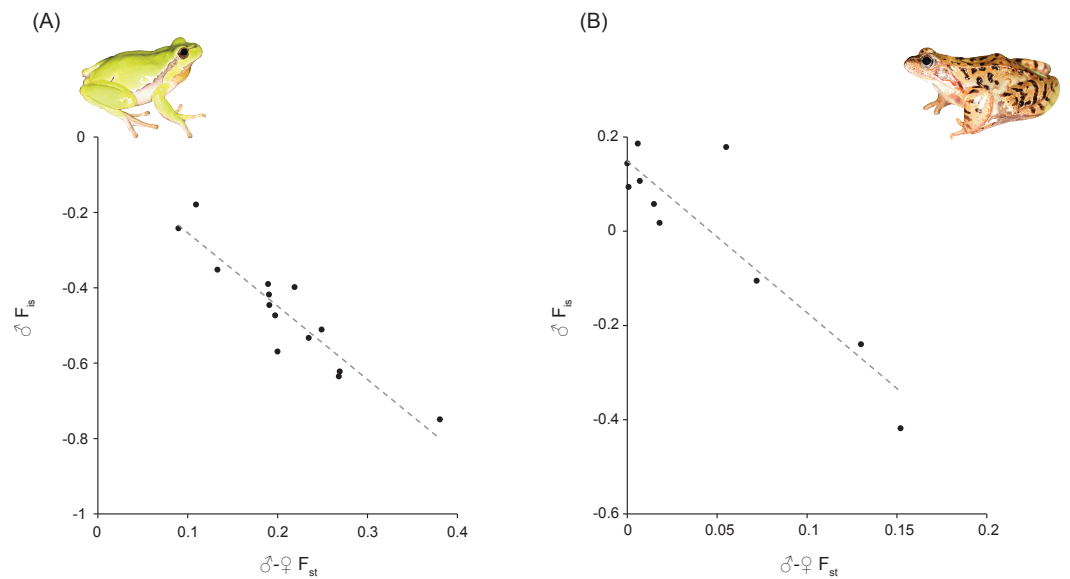
Figure 1 F_{st} **between sexes ($\male$–$\female F_{st}$) versus male F_{is} ($\male F_{is}$) at sex-linked loci in *Hyla arborea* and *Rana temporaria*.** Both are highly significant (Table 1). Photo credit: Christophe Dufresnes.

Table 1 **Correlation between male F_{is} ($\male F_{is}$), F_{st} between sexes ($\male$–$\female F_{st}$) and X–Y differentiation (X–Y dif.) at sex-linked loci.** $\male F_{is}$ was also adjusted by F_{is} at autosomal loci (auto. F_{is}) and F_{is} at sex-linked loci in female ($\female F_{is}$).

	H. arborea			R. temporaria		
	N	*R²*	*P*	*N*	*R²*	*P*
$\male F_{is}$ *vs.* $\male$–$\female F_{st}$	14	0.86	<0.001	10	0.82	<0.001
$\male F_{is}$ (adjusted by auto. F_{is}) *vs.* $\male$–$\female F_{st}$	14	0.86	<0.001	–	–	–
$\male F_{is}$ (adjusted by $\female F_{is}$) *vs.* $\male$–$\female F_{st}$	14	0.70	<0.001	10	0.90	<0.001
$\male$–$\female F_{st}$ *vs.* X–Y dif.	14	0.71	<0.001	–	–	–
$\male F_{is}$ *vs.* X–Y dif.	28	0.75	<0.001	–	–	–
$\male F_{is}$ (adjusted by auto. F_{is}) *vs.* X–Y dif.	27	0.70	<0.001	–	–	–
$\male F_{is}$ (adjusted by $\female F_{is}$) *vs.* X–Y dif.	14	0.43	0.010	–	–	–

Notes.
Abbreviations: N, number of populations; R^2, fit of linear regression; P, p-value of linear regressions.

RESULTS & DISCUSSION

We established significant correlations between the different statistics for both species (Fig. 1 and Table 1). As expected, $\male F_{is}$ is negatively correlated with F_{st} between sexes (for *H. arborea*: $R^2 = 0.86$; for *R. temporaria*: $R^2 = 0.82$). Moreover, for *H. arborea*, we can further show that these two estimates are well-correlated with a measure of X–Y differentiation computed from phased genotypes (for $\male F_{is}$: $R^2 = 0.75$; for $\male$–$\female F_{st}$: $R^2 = 0.71$; Fig. 2 and Table 1). Thus, both statistics appear as reliable proxies to estimate overall differentiation between sex chromosomes.

However, we further report strong variation among the individual fits of each locus in both species (Figs. S1 and S2). The R^2 associated with the regressions of $\male F_{is}$ by
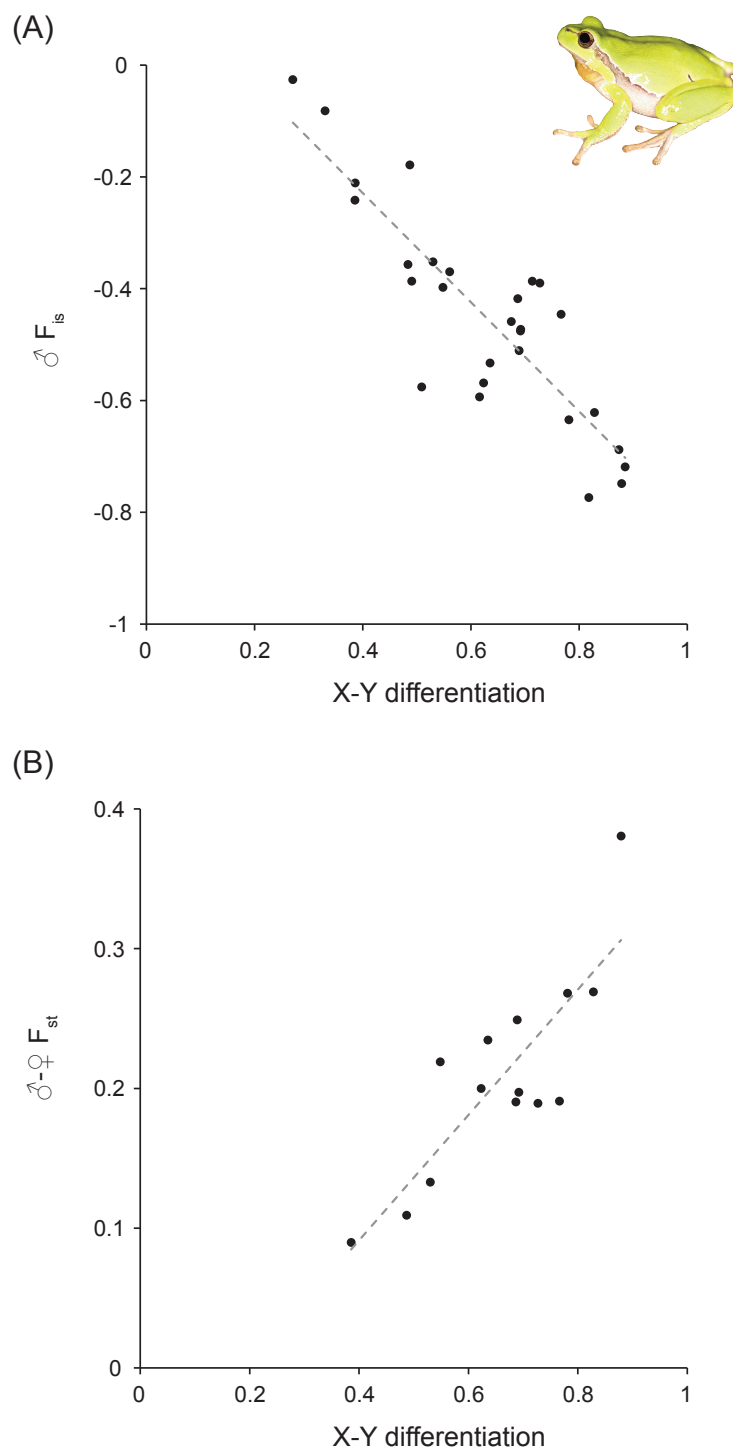
Figure 2 **X–Y differentiation versus male F_{is} ($\male F_{is}$) and F_{st} between sexes ($\male-\female F_{st}$) at sex-linked loci in** *Hyla arborea.* Both are highly significant (Table 1). Photo credit: Christophe Dufresnes.

$\male$–$\female F_{st}$ averaged 0.54 ± 0.32 for *H. arborea* (Fig. S1) and 0.57 ± 0.33 for *R. temporaria* (Fig. S2). Although lower sample sizes may account for part of this variation (as some loci were not informative in every populations), such fluctuations may also likely be due by stochastic processes like drift. Thus, at least several markers appear needed to obtain sound estimations. While this is usually the case for studies of whole-chromosome differentiation (e.g., *Dufresnes et al., 2014a*; *Dufresnes et al., 2014b*), it might become an issue for comparing fine-scale patterns along chromosomal segments (e.g., sliding window analyses), which then requires a denser coverage to obtain meaningful estimates.

The $\male F_{is}$ statistic is also expected to be affected by the baseline level of inbreeding in populations. Here it should not have impacted the comparisons for *H. arborea*, since the populations analyzed are known to meet Hardy–Weinberg Equilibrium (HWE), as inferred from autosomal markers (*Dufresnes et al., 2013*). Accordingly, controlling $\male F_{is}$ by autosomal F_{is} yielded similarly good correlations (Table 1, Fig. S1). In parallel, we also tested whether F_{is} at sex-linked markers in females ($\female F_{is}$) could be used for the adjustments instead, in absence of autosomal data. The resulting fits were quite variable, being overall better for *R. temporaria*, but worse for *H. arborea* (Table 1, Figs. S1 and S2). These inconsistencies may indicate that $\female F_{is}$ is a poor corrector for such analysis. One explanation probably lies within the effective size of X chromosomes, which depends on their amount of recombination with the Y, i.e., $\frac{3}{4}$ of autosomes if X–Y recombination is suppressed, but similar to autosomes if both copies freely recombine. Here it should strongly fluctuate among the different populations considered, given their contrasted sex-chromosome dynamics. In *H. arborea*, X–Y recombination rates were shown to evolve rapidly and strongly vary between populations (*Dufresnes et al., 2014a*; *Dufresnes et al., 2014b*). In *R. temporaria*, sex-determination is not strictly genetic, and so the same loci behave either like non-recombining sex chromosomes, or autosomes, depending on populations (*Rodrigues et al., 2014*; *Rodrigues et al., 2015*; *Rodrigues et al., 2016*). In parallel, sex-biased dispersal may also account for such discrepancies, by inflating F_{is} of the dispersing sex (i.e., towards a larger heterozygote deficit, *Goudet, Perrin & Waser, 2002*). Some evidence did suggest sex-biased dispersal in our focal species, i.e., male-biased in *H. arborea* (based on capture-mark-recapture data; *Vos, Ter Braak & Nieuwenhuizen, 2000*) but female-biased in *R. temporaria* (based on genetic data; *Palo et al., 2004*). Therefore, given our results and the potential cofounding factors affecting sex-specific F_{is}, autosomal F_{is} (ideally computed from samples of both sexes) should thus rather be considered to correct sex-linked $\male F_{is}$, whenever possible. Moreover, allele dropout, which is inherent to some commonly used genotyping-by-sequencing methods like RAD (Restriction site-associated DNA), can lead to overestimate F_{is} (*Gautier et al., 2013*). However, this process being likely random, it should similarly affect autosomal and sex-linked markers; $\male F_{is}$ relative to autosomal F_{is} should thus be comparable among populations.

The low sampling requirement for computing these F-statistics significantly simplifies population genetic analyses of homomorphic sex-chromosomes. F_{st} between sexes was used to this purpose in our previous studies to investigate the geographic patterns of sex-chromosome differentiation (*Rodrigues et al., 2013*; *Rodrigues et al., 2014*; *Dufresnes et al., 2014b*), with coherent results. Moreover, sex-linked $\male F_{is}$, was also successfully applied in

studies of sex-chromosome differentiation in stickleback fishes (*Shikano et al., 2011*; *Natri, Shikano & Merilä, 2013*). Importantly, $\male F_{is}$ has the advantage not to rely on female genotypes, which are usually the conspicuous sex and are thus harder to sample in many species. This metric actually opens opportunities to exploit sample series that were not originally designed for sex-chromosome studies (e.g., museum collections), and where a majority of males is represented. Furthermore, these approaches should also be applicable to female-heterogametic systems (ZW), by computing $\female F_{is}$. In fact, due to the high recombination rates usually observed in females (*Brelsford, Dufresnes & Perrin, 2016a*; *Brelsford, Rodrigues & Perrin, 2016*), reconstructing Z and W haplotypes may be virtually impossible, so $\female F_{is}$ and $\male - \female F_{st}$ would be the only way to compare Z–W differentiation between populations. Combining these simple statistics with population genomic data will guarantee exciting new insights into the unusual ways sex chromosomes evolve in many organisms.

ACKNOWLEDGEMENTS

We thank G Lavanchy and D Jeffries for insightful discussions and N Perrin for support, as well as the editor and two anonymous reviewers for their useful comments.

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Nicolas Rodrigues analyzed the data, prepared figures and/or tables, reviewed drafts of the paper.
- Christophe Dufresnes analyzed the data, wrote the paper, prepared figures and/or tables.

Data Availability

The raw data originates from published studies. Statistics used in this study are available as Table S1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.3207#supplemental-information.

REFERENCES

Beukeboom LW, Perrin N. 2014. *The evolution of sex determination.* Oxford: Oxford University Press.

Brelsford A, Dufresnes C, Perrin N. 2016a. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity* **116**:177–181 DOI 10.1038/hdy.2015.83.

Brelsford A, Dufresnes C, Perrin N. 2016b. Trans-species variation in *Dmrt1* is associated with sex determination in four European tree-frog species. *Evolution* **70**:840–847 DOI 10.1111/evo.12891.

Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N. 2017. Identifying homomorphic sex chromosomes from wild caught adults with limited genomic resources. *Molecular Ecology Resources* In Press.

Brelsford A, Rodrigues N, Perrin N. 2016. High-density linkage maps fail to detect any genetic component to sex determination in a *Rana temporaria* family. *Journal of Evolutionary Biology* **29**:220–225 DOI 10.1111/jeb.12747.

Dufresnes C, Bertholet Y, Wassef J, Ghali K, Savary R, Pasteur B, Brelsford A, Rozenblut-Koscisty B, Ogielska M, Stöck M, Perrin N. 2014b. Sex-chromosome differentiation parallels postglacial range expansion in European tree frogs (*Hyla arborea*). *Evolution* **68**:3445–4556 DOI 10.1111/evo.12525.

Dufresnes C, Borzée A, Horn A, Stöck M, Ostini M, Sermier R, Wassef J, Litvinchuck SN, Kosch TA, Waldman B, Jang Y, Brelsford A, Perrin N. 2015. Sex-chromosome homomorphy in Palearctic tree frogs results from both turnovers and X–Y recombination. *Molecular Biology and Evolution* **32**:2328–2337 DOI 10.1093/molbev/msv113.

Dufresnes C, Stöck M, Brelsford A, Perrin N. 2014a. Range-wide sex-chromosome sequence similarity supports occasional XY recombination in European tree frogs (*Hyla arborea*). *PLOS ONE* **9**:e97959 DOI 10.1371/journal.pone.0097959.

Dufresnes C, Wassef J, Ghali K, Brelsford A, Stöck M, Lymberakis P, Crnobrnja-Isailovi J, Perrin N. 2013. Conservation phylogeography: does historical diversity contribute to regional vulnerability in European tree frog (*Hyla arborea*). *Molecular Ecology* **22**:5669–5684.

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**:3165–3178 DOI 10.1111/mec.12089.

Goudet J. 1995. FSTAT (version 1.2): a computer program to calculate *F*-Statistics. *Journal of Heredity* **86**:485–486 DOI 10.1093/oxfordjournals.jhered.a111627.

Goudet J, Perrin N, Waser P. 2002. Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Molecular Ecology* **11**:1103–1114 DOI 10.1046/j.1365-294X.2002.01496.x.

Kirkpatrick M, Guerrero RF. 2014. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* **197**:2531–2541 DOI 10.1534/genetics.113.156026.

Natri MN, Shikano T, Merilä J. 2013. Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. *Molecular Biology and Evolution* **30**:1131–1144 DOI 10.1093/molbev/mst035.

Palo JU, Lesbarrères D, Schmeller DS, Primmer CR, Merilä J. 2004. Microsatellite marker data suggest sex-biased dispersal in the common frog *Rana temporaria*. *Molecular Ecology* **13**:2865–2869 DOI 10.1111/j.1365-294X.2004.02286.x.

Perrin N. 2016. Random sex determination: when developmental noise tips the sex balance. Perrin N. *BioEssays* **38**:1218–1226 DOI 10.1002/bies.201600093.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Rodrigues N, Betto-Colliard C, Jourdan-Pineau H, Perrin N. 2013. Within-population polymorphism of sex-determination systems in the common frog (*Rana temporaria*). *Journal of Evolutionary Biology* **26**:1569–1577.

Rodrigues N, Merilä J, Patrelle C, Perrin N. 2014. Geographic variation in sex-chromosome differentiation in the common frog (*Rana temporaria*). *Molecular Ecology* **23**:3409–3418 DOI 10.1111/mec.12829.

Rodrigues N, Vuille Y, Brelsford A, Merilä J, Perrin N. 2016. The genetic contribution to sex determination and number of sex chromosomes vary among populations of common frogs (*Rana temporaria*). *Heredity* **117**:25–32 DOI 10.1038/hdy.2016.22.

Rodrigues N, Vuille Y, Loman J, Perrin N. 2015. Sex-chromosome differentiation and 'sex races' in the common frog (*Rana temporaria*). *Proceedings of the Royal Society of London B* **282**:20142726 DOI 10.1098/rspb.2014.2726.

Shikano T, Natri MN, Shimada Y, Merilä J. 2011. High degree of sex chromosome differentiation in stickleback fishes. *BMC Genomics* **12**:474 DOI 10.1186/1471-2164-12-474.

Stöck M, Savary R, Betto-Colliard C, Biollay S, Jourdan-Pineau H, Perrin N. 2013. Low rates of X–Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). *Journal of Evolutionary Biology* **26**:674–682 DOI 10.1111/jeb.12086.

Vos CC, Ter Braak CJF, Nieuwenhuizen W. 2000. Incidence function modelling and conservation of the tree frog *Hyla arborea* in the Netherlands. *Ecological Bulletins* **48**:165–180.

Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. *Nature Communication* 12087.

Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A, Fujiyama A, Kitano J. 2014. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLOS Genetics* **10**:e1004223 DOI 10.1371/journal.pgen.1004223.