DR. HANNA LAAKKONEN (Orcid ID : 0000-0003-1762-4509)

DR. DANIELE SALVI (Orcid ID : 0000-0002-3804-2690)

Article type : Original Article

Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard

Running title: Asymmetric introgression in wall lizards

Weizhao Yang[1], Geoffrey M While[2], Hanna Laakkonen[1], Roberto Sacchi[3], Marco A L Zuffi[4], Stefano Scali[5], Daniele Salvi[6,7]and Tobias Uller[1,8]

1. Department of Biology, Lund University
2. School of Biology, University of Tasmania
3. Department of Earth and Environmental Sciences, University of Pavia
4. Museum Natural History, University of Pisa
5. Museum of Natural History of Milan
6. Department of Health, Life and Environmental Sciences, University of L'Aquila; 67100 Coppito, L'Aquila, Italy
7. CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos; University of Porto; Campus Agrário de Vairão, 4485-661 Vairão, Portugal
8. Edward Grey Institute, Department of Zoology, University of Oxford

Corresponding author: Tobias Uller, tobias.uller@biol.lu.se

Abstract

Strongly selected characters can be transferred from one lineage to another with limited genetic exchange, resulting in asymmetric introgression and a mosaic genome in the receiving population. However, systems are rarely sufficiently well studied to link the pattern of introgression to its underlying process. Male common wall lizards in western Italy exhibit exaggeration of a suite of sexually selected characters that make them outcompete males from a distantly related lineage that lack these characters. This results in asymmetric hybridization and adaptive introgression of the suite of characters following secondary contact. We developed genome-wide markers to infer the demographic history of gene flow between different genetic lineages, identify the spread of the sexually selected syndrome, and test the prediction that introgression should be asymmetric and heterogeneous across the genome. Our results show that secondary contact was accompanied by gene flow in both directions across most of the genome, but with approximately 3% of the genome showing highly asymmetric introgression in the predicted direction. Demographic simulations reveal that this asymmetric gene flow is more recent than the initial secondary contact, and the data suggest that the exaggerated male sexual characters originated within the Italian lineage and subsequently spread throughout this lineage before eventually reaching the contact zone. These results demonstrate that sexual selection can cause a suite of characters to spread throughout both closely and distantly related lineages with limited gene flow across the genome at large.

Keywords: asymmetric introgression, sexual selection, mosaic genome, *Podarcis muralis*

Introduction

A major goal of ecology and evolution is to understand how historical processes gave rise to contemporary patterns of diversity. The recent proliferation of genomic data has revealed that phenotypes that evolved in one lineage can sometimes spread to other lineages with limited genetic exchange, causing mosaic genomes and a mismatch between phenotypic and genetic variation (reviewed in Palmer& Kronforst 2015; Arnold 2015).When introgression involves phenotypes and genotypes that have a selective advantage, this may enable more rapid evolution than possible with other sources of genetic variation (Hedrick 2013; Arnold & Kunte 2017). For example, introgression between lineages appears to have facilitated the evolution of mimetic wing pattern and colouration in *Heliconius* butterflies (Heliconius Genome Consortium 2012). Such signatures of potentially adaptive introgression from genome scans have now been found in a variety of plants and animals, including humans (e.g., Song et al. 2011; Huerta-Sanchez et al. 2014; Lamichhaney et al. 2015).

Studies of adaptive introgression are empowered by an understanding of what makes individuals with particular traits have higher survival or reproductive success. One such example is the adaptive transfer of a suite of sexually selected characters across a contact zone between two genetically divergent lineages of the common wall lizard, *Podarcis muralis* (While et al. 2015). The suite of characters – or syndrome – includes large body and head size, exaggeration of visual signals, in particular colouration, and aggressive behaviours that give males a competitive advantage over males with the ancestral wall lizard phenotype (Heathcote et al. 2016; MacGregor et al. 2017a; Fig. 1a,b). This advantage results in asymmetric hybridization upon secondary contact, which is consistent with the extensive introgression of the sexually selected characters that is observed in both the native contact zone, situated along the north-western coast of Italy, and in non-native populations in England and Germany (While et al. 2015). As a result of this introgression, lizards in the

native hybrid zone exhibit a conspicuous mismatch between the expression of sexually selected characters and patterns of genetic variation measured with mitochondrial and microsatellite genetic markers, with genetic markers being representative of a distantly related lineage (the 'Southern Alps' lineage; While et al. 2015; see Schulte et al. 2012 and Salvi et al. 2013 for phylogeography).

While it is well understood that differences in male competitive ability drive the asymmetric introgression of sexually selected traits in common wall lizards, it is unknown to what extent their spread is accompanied by asymmetric introgression across the genome. On the one hand, the relevant phenotypes, including morphology and colouration, are quantitative traits and hybrids can express a continuum of intermediate phenotypes. This implies that lineage divergence should be associated with divergence across many loci and hence that introgression of the suite of characters should be accompanied by substantial genetic exchange across the genome. On the other hand, the characters are highly correlated both within populations of Italian origin (MacGregor et al. 2017a) and across the hybrid zone (While et al. 2015). For example, the geographic clines for dorsal colouration, ventral colouration and relative head size are highly concordant (While et al. 2015). This suggests that the characters are functionally and genetically integrated, which may result in adaptive introgression with limited genetic exchange overall.

Identifying the extent to which phenotypic introgression is accompanied by genetic exchange is an important step for the interpretation of the historical processes that may have led to these patterns as well as their evolutionary consequences. To address this, we developed genome-wide genetic markers to (i) study the relationship between phenotypes and population structure in the relevant geographic region, (ii) infer the demographic history of gene flow between genetic lineages, and (iii) test the prediction that genetic introgression should be asymmetric and heterogeneous across the genome.

Methods

The study system

The common wall lizard (*Podarcis muralis*) is a small diurnal lizard that is wide-spread in Mediterranean and temperate Europe. It is sexually dimorphic throughout its range, but sexually selected characters are particularly striking in central Italy (Latium and Tuscany). These characters include a large head size and body mass relative to skeletal size and conspicuous colouration, in particular more highly expressed lateral ocelli, extensive black colouration of the throat, chest and belly and green dorsal colouration. Individuals elsewhere in the species' distribution exhibit smaller ocelli, very limited blackness on ventral scales, and brown dorsal colouration. Populations with exaggerated sexual characters have been designated to the subspecies *Podarcis muralis nigriventris* (Bonaparte 1836, considered by Gruschwitz and Böhme 1986; Böhme 1986), but this taxonomic unit is not supported genetically (see Results). Morphology and colouration are positively correlated with aggressive behaviour and dominance, and individuals with more exaggerated phenotypes have higher reproductive success, both in competition with males from the same region (MacGregor et al. 2017a) and in competition with males from western Europe with the ancestral phenotype (While et al. 2015; Heathcote et al. 2016; MacGregor et al. 2017a). Females do not discriminate between males of different origins and there is no female choice on male quantitative characters (While et al. 2015; Heathcote et al. 2016; While & Uller 2017). As a consequence, hybridization in experimental contact zones mainly involves males of the dominant lineage, despite that males prefer to court females that belong to their own genetic lineage (Heathcote et al. 2016; MacGregor et al. 2017a). The phenotypic differences between lizards with the sexually selected syndrome and lizards with the ancestral phenotype persists in non-native populations in England and in captivity (e.g., While et al. 2015;

MacGregor et al. 2017c), strongly suggesting that the influence of phenotypic plasticity on lineage differences is minor.

We have previously shown (While et al. 2015) that there is a hybrid zone along the north-western coast of Italy (Liguria) between the genetic lineage predominantly found in Latium and Tuscany (here referred to as the 'Italian lineage' for short, but lizards from this region are also sometimes described as the 'Tuscan lineage' based on mitochondrial haplotypes, e.g., Schulte et al. 2012) and a genetic lineage found in north-western Italy (the 'Southern Alps' lineage; While et al. 2015; Fig.1c). In alignment with the experimental data reviewed above, there is substantial asymmetric phenotypic introgression from the Italian lineage into the Southern Alps lineage. Indeed, lizards from Southern Alps populations that do not show any detectable genetic evidence of introgression across mtDNA and microsatellite markers can be phenotypically nearly indistinguishable from lizards belonging to the Italian genetic lineage that exhibit exaggerated sexually selected characters (While et al. 2015).

Field sampling

Individuals for the present study were sampled from 45 locations in central and western Italy (Fig.1c and Table S1). This covered the previously established hybrid zone (While et al. 2015) as well as neighbouring regions where lizards exhibit the suite of sexually selected characters (e.g., intense green dorsal colouration, extensive black ventral colouration, large body and head size), the ancestral phenotype, or different degrees of intermediate phenotypes. Note that the populations in the south-west of our sampling design (Fig. 1c) represent the southernmost contemporary range of this lineage and that the species distribution is discontinuous and restricted to the Apennine mountains in southern Italy (e.g., Sindaco et al., 2006; Salvi et al. 2013).

Full details of the field methods and the phenotypic measurements can be found in While et al. (2015). In brief, at each location, we captured ≥ 25 lizards with approximately equal sex ratios by noosing. Upon capture we measured their morphological traits (snout-to-vent length (SVL), total length, head length, head width, body mass), and photographed them using standard methods (see While et al. 2015 for extensive details). Sample collection was carried out in accordance with local laws and regulations in Italy, under the collection permit number Prot. PNM-2015-0009720.

We scored the intensity of the dorsal colouration ('greenness') and the extent of black ventral coloration ('blackness'). Dorsal greenness and blackness are useful markers of phenotypic introgression since they are highly divergent between lineages and correlate with other sexually selected characters (While et al. 2015; MacGregor et al. 2017a, 2017b). More extensive cline analysis of phenotypes in both males and females can be found in While et al. (2015) and MacGregor et al. (2017b). The intensity of dorsal greenness is scored on a scale from one to ten, which is highly repeatable and strongly correlated with objective spectrophotometry (While et al. 2015). Blackness is scored by the proportion of black colour on the chest area and thus ranges from zero to one. Tissue samples from the tail or toe (in case of tail loss) were collected for DNA extraction. All lizards were released at the location of capture after processing.

Sequencing and genotyping

From each location, up to 7 males and 7 females (all adults, SVL >48 mm) were selected randomly with respect to phenotype for sequencing and genotyping, resulting in a total of 597 individuals from 45 locations (see Table S1for sample sizes per location). Genomic DNA was extracted using DNEasy blood and tissue kit (Qiagen, USA). We generated double-digest restriction site-associated DNA (ddRAD) markers following the protocol outlined by

Peterson et al. (2012). DNA was digested with the restriction enzymes EcoRI High Fidelity and MspI (New England Biolabs, USA). For each individual, we used 500 ng of DNA at a standardized concentration greater than 20ng/uL (quantified using a Qubit fluorometer; Life Technologies, USA). We selected fragments of a size between 300 and 700 bp, and amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA) with 12 cycles. Finally, the libraries were sequenced paired-end with read length of 150 bp on an Illumina HiSeq 2500 platform in Novogene Company Limited (Hong Kong).

STACKS (v1.47, Catchen et al. 2011) was used to process the sequence reads and call single nucleotide polymorphisms (SNPs) for each individual. First, 'process_radtags' module was used to filter sequence reads with low quality score (Phred score threshold was set to 30), ambiguous base call, and incomplete barcode or restriction site. Clean reads were mapped to a draft genome of common wall lizard (unpublished) using GSNAP (v2017-05-03, Wu et al. 2016). The draft genome was sequenced from a male individual from the Iberian Peninsula, using both Illumina and PacBio platforms. The assembly is a total of 1.51 Gb with 2162 contigs and the N50 size is 92.4 Mb. Alignment bam files were used as input for the reference-based STACKS pipeline 'ref_map.pl' to estimate the effective RAD sequence tags and detect SNPs for each sample under SNP model with p-value 0.05.

VCFTools (v0.1.14, Danecek et al. 2011) was used to calculate the distribution of average depth of coverage per SNP across all samples. SNPs with depth < 5^{th} percentile of the distribution (12) were removed as low-depth loci. In addition, we also filtered the SNPs with depth greater than 95^{th} percentile (62) to remove SNPs affected by PCR duplicates or SNPs from high complexity regions (Fan et al. 2014). The individuals with average depth of coverage for all SNPs less than 12 were also removed from the whole dataset. Plink (v1.9, Chang et al. 2015) was used to filter the SNPs with minimum minor allele frequency (MAF) less than 0.05, deviation from Hardy-Weinberg Equilibrium (P<0.05), or with missing rate

greater than 0.1 in each population. Individuals with genotyping rate < 90% were also excluded. Finally, for population structure analysis we used Plink (v1.9,Chang et al. 2015) to filter the SNPs in linkage disequilibrium with correlation coefficient $r^2 > 0.2$ (the parameter is: --indep-pairwise 50 5 0.2).

Population structure analysis

To infer the genetic relationship for all populations, we used three approaches to study their population structure. Principal component analysis (PCA) was first performed using Plink for all populations. ADMIXTURE (v1.3.0, Alexander et al. 2009) was used to estimate individual admixture assuming different numbers of clusters with the co-ancestry clusters (K) spanning from 2 to 15. Cross-validation (CV) was set to 10 folds to compare different number of K, in which lower CV value indicates the most likely number of clusters. We also constructed the neighbour-joining tree based on pairwise genetic distance matrix data of all individuals calculated by Neighbor in Phylip (v3.697, Felsenstein 2005).

Geographic cline analysis

Mitochondrial DNA and microsatellite markers have revealed a hybrid zone between the Italian and Southern Alps lineages along the Mediterranean coast (While et al. 2015). Here we revisited this geographic cline, adding two additional locations at the center of the hybrid zone, to establish the patterns of introgression across the genome. We used the probability that an individual was assigned by ADMIXTURE (v1.3.0, Alexander et al. 2009) to Italian lineage (Q) as the hybrid index. An individual was classified as pure Italian or Southern Alps individual if $Q \geq 0.9$ or $Q \leq 0.1$, and mixed ancestry if $0.9 > Q > 0.1$. The hybrid indices for 13 populations from Loano (LO) to Chianni (CN) in a line transect across the hybrid zone (calculated by point-to-point distance; While et al. 2015) was fitted to a series of equilibrium

geographic cline models using the Metropolis-Hastings Markov chain Monte Carlo algorithm employed in the R package HZAR (v0.2-5, Derryberry et al. 2014). We ran 15 separate models that varied in the number of cline shape parameters estimated for hybrid indices and selected the model with the lowest Akaike information criterion (AIC) as the best fitting cline (Derryberry et al. 2014). We also fitted greenness and blackness, two representative traits of the syndrome, to phenotype cline models in the same transect for males, females and both sexes, by using five separate models designed for phenotypic data (Derryberry et al. 2014). Geographic cline for all the candidate loci for the sexually selected phenotype ('candidate loci index';see below) was also fitted using the same approach as hybrid index. In addition, individual SNPs from the candidate list with frequency >0.75 on one end and <0.25 on the other end of the cline were also fitted to investigate their cline center or width (Baldassarre et al. 2014). The coincidence of cline centers for hybrid index, candidate loci, and representative phenotypes was estimated using the maximum-likelihood derived confidence intervals (While et al. 2015).

Detection and characterization of loci associated with the sexually selected phenotype

To study the genomic signatures of sexually selected introgression in *P. muralis*, we identified candidate loci across the genome for individuals with the sexually selected phenotype. As characters are highly correlated, we used a cutoff corresponding to a greenness score of eight or higher to identify the most extreme populations from Latium and Tuscany as a Syndrome Group (Latium, LA-S: FO, FU, RO, and SMA; Tuscany, TU-S: CL, CN, and CR; see While et al. 2015 for colour score). Individuals with an equally highly expressed phenotype but clustering within the Southern Alps lineage (populations, SA-S: GN, RA, and ST) were also included to reveal the introgression across genetic lineages. Neighbouring populations with minimal expression of the syndrome (corresponding to a greenness score of

one or two) were grouped as the Reference Group (Latium, LA-R: DS, FG, and OV; East Tuscany, TU-S: BE and BL; North Tuscany, TUN-R: MP and MV; Southern Alps, SA-R: LO, NL, and SO).

Two complementary approaches were used to identify genomic regions potentially affected by selection on the syndrome, both of which were based on allele frequencies of variable sites. We first calculated the sequence diversity statistics (π) by using VCFTools (v0.1.14, Danecek et al. 2011) for each RAD sequence contig in the Syndrome and Reference Groups, respectively. The log value of the ratio in Reference Group to that in the Syndrome Group (π_R/π_S) was used to detect the signature of selection in Syndrome Group (Huang et al. 2012). We also calculated population fixation statistics (F_{ST}) between the two groups and Z-transformed the value. The putative selection targets were extracted based on being in the top 5% of log-odds ratios for both π and F_{ST} (Chen et al. 2016) (Fig. 2a). The genetic differentiation patterns were compared between candidate loci and non-candidate ('neutral') loci using pairwise F_{ST} analysis for populations from different regions (Izuno et al. 2017). To further study the genetic differentiation of the candidate loci , we conducted the neighbour-joining tree analysis for individuals from the Syndrome and Reference Groups. We limited this analysis to the Italian lineage, which avoids the influence of hybridization between Italian and Southern Alps lineage against the tree topology. We conducted the neighbour-joining tree analysis by TreeBeST (Viella et al. 2009) with 1000 times of bootstrap for both candidate loci and neutral loci, respectively.

Demographic simulation of gene flow between the two genetic lineages

To reveal the population dynamics and how gene flow varied in history, the demographic patterns of Italian and Southern Alps lineages were inferred using a modified version of the software δaδi (v1.7, Gutenkunst et al. 2009) based on the joint allele frequency spectrum

(JAFS) of SNPs (excluding populations MG and VI; Fig. 1d). In δaδi, each demographic model consists of a series of population parameters, including population size (N), time scale (T), and possibly exchanging migrants (m). We performed two rounds of simulation. In the first round, four basic models were fitted representing alternative modes of divergence between the two genetic lineages: Strict Isolation (SI), Isolation-with-Migration (IM), Ancient Migration (AM), and Secondary Contact (SC) (Fig. 3a). In the second round, models were extended to consider heterogeneous migration across the genome. In order to quantify the mean effect of selection at specific sites, Tine et al. 2014 and Rougeux et al. 2017 defined a Hill-Robertson scaling factor (hrf), relating the effective population size of loci influenced by selection to that of neutral loci. Those parameters were implemented in the best-fitting basic model (i.e. SC) in two different scenarios as SC2M-1 and SC2M-2. In SC2M-1, heterogeneous gene flow was set to occur concurrently with the time of the secondary contact; whereas in SC2M-2, heterogeneous migration was set to occur later than the secondary contact (Fig. 3b). Due to the lack of reference genome of a related species as outgroup, SNP loci were folded in the simulations. Models were fitted independently using successively a hot and a cold simulated annealing procedure followed by "BFGS" optimization (Gutenkunst et al. 2009). We ran 20 independent optimizations for each model in order to check for convergence and retained the best one for comparisons among models based on Akaike information criterion (AIC).

Results

In total, we obtained RAD-seq data for 597 common wall lizards from 45 populations (Fig.1c and Table S1). RAD sequencing generated a total of 4.1 billion paired-end reads of 150 bp read length. On average, 73.11% of the reads could be mapped to the reference genome for all samples, ranging from 63.77%-82.48%. After stringent quality filtering, 589 individuals

were retained and genotyped at 41,543 single nucleotide polymorphisms (SNPs) resulting from 20,015 ddRAD sequence contigs. The mean coverage depth for RAD contigs was 31X per site and the average genotyping rate was 97.3%. Of these, we obtained 22,446 unlinked SNPs for population structure analysis after removing sites under linkage disequilibrium.

Population structure and genetic relationship

In a principal component analysis (PCA), the first principal component (variance explained = 58.4%) separated all populations into two major groups: the populations along the north-western coast (referred to as the Southern Alps lineage), and all populations in central Italy (referred to as the Italian lineage). The second principal component (variance explained = 9.6%) indicated divergence within the Italian lineage from south to north. Two populations, MG and VI, showed clear evidence for the presence of hybrids between the Southern Alps and Italian lineages (Fig.1c,d), which was also supported by the hybrid index (Q) that revealed mixed ancestry for MG and VI (hybrid index Q_{MG}=0.31 and Q_{VI}=0.84 respectively).Clustering analysis corroborated the separation between Southern Alps and Italian lineages when the number of presumed ancestral population (K) was set to two (CV=0.471). The Italian lineage was subdivided into north and south populations with K=3 (CV=0.463). When K=4 (CV=0.458) and K=5 (CV=0.454), populations in Tuscany were further subdivided into three clusters; center, east, and north (Fig.1e). A good correspondence was also found with results from the neighbour-joining tree, where the Southern Alps and Italian lineages were clearly separated (Fig.1f). It is worth noting that the Italian lineage includes not only the populations with the sexually selected phenotype (around Latium and Tuscany), which had been historically assigned to '*P. m. nigriventris*' subspecies, but also the other populations without this phenotype. Thus, the subspecies '*P. m. nigriventris*' is not a valid taxonomic unit.

Geographic cline analysis

The best-fitting cline center for the hybrid index created from the SNP data was 162.20 (157.8-185.3) km from the westernmost population (LO). As predicted, and consistent with previous results (While et al. 2015), the best-fitting cline centers for both greenness (c=47.16 (34.7-78.4) km from LO) and blackness (c=62.8 (9.4-89.1) km from LO) were shifted further to the west than the genetic cline (Fig.1g, and Table S4). Clines for the representative phenotypes fitted separately for males and females showed a similar pattern (Fig. S1 and Table S4).

Detection and characterization of loci associated with the syndrome

In total, we identified 438 outlier sequence contigs with a selective signature in the Syndrome Group compared to the Reference Group, accounting for 2.19% of all the RAD contigs (Fig.2a). The genetic differentiation of the candidate loci among populations was quite different from that of the neutral loci. The mean pair-wise F_{ST} for candidate loci of neighbouring populations from the Syndrome Group ($F_{ST(TU\text{-}S,\ LA\text{-}S)}$=0.032; $F_{ST(SA\text{-}S,\ TU\text{-}S)}$=0.182; subscripts stand for the main genetic clusters; see Methods and Fig. 2b) was lower than that for neutral loci ($F_{ST(TU\text{-}S,\ LA\text{-}S)}$=0.054; $F_{ST(SA\text{-}S,\ TU\text{-}S)}$=0.358) (Fig. 2b). However, F_{ST} values for candidate loci showed a striking increase between populations from the Syndrome Group and neighbouring populations of the Reference Group, especially in Tuscany (candidate loci: $F_{ST(TU\text{-}S,\ TUE\text{-}R)}$=0.209, $F_{ST(TU\text{-}S,\ TUN\text{-}R)}$=0.184; neutral loci: $F_{ST(TU\text{-}S,\ TUE\text{-}R)}$=0.069, $F_{ST(TU\text{-}S,\ TUN\text{-}R)}$=0.059) and Southern Alps (candidate loci: $F_{ST(SA\text{-}S,\ SA\text{-}R)}$=0.307; neutral loci: $F_{ST(SA\text{-}S,\ SA\text{-}R)}$=0.078). In addition, the best-fitting cline for the candidate loci index across Italian and Southern Alps lineages had its center 68.54 (52.66-80.79) km from LO, coinciding with both greenness and blackness, but deviating from the hybrid index (Fig. 1g, and Table S4). Among 117 SNPs from the candidate list with frequency > 0.75 and <

0.25 on two ends of the cline, 98 deviated from the cline for hybrid index, and 82 coincided with the clines for the representative phenotypes (Fig. S2). Those results are indicative of a mosaic genome and possible introgression of candidate loci into the main genetic clusters (Fig.2b).

To further investigate the genetic differentiation of the candidate loci, we constructed the neighbour-joining tree for individuals from both the Syndrome and the Reference groups with in Italian lineage. The topology of the tree based on neutral loci showed that the individuals from Latium and Tuscany still clustered into two independent clades no matter what their phenotypes were (Fig. 2c). In contrast, for the candidate loci, all individuals from the Syndrome groups clustered together in a monophyletic clade with a high supporting rate (96). Individuals from LA-R group were positioned at the basal position together with the Syndrome clade, while the supporting rate for the node was not high (25, Fig. 2d). In addition, TU-S clade clustered with Latium populations and departed from other Tuscan clade (TUE-R and TUN-R) in the candidate loci tree, which was consistent with the result of pairwise F_{ST} scores (Fig. 2b, d). This implies that the candidate alleles for the syndrome might have originated from the south, around Latium, and then expanded into Tuscany.

Demographic simulation of gene flow between the two genetic lineages

To infer the history of gene flow between the two main lineages we ran a series of demographic models. In the first round of simulation, four basic models (SI, IM, AM, and SC) were fitted to reveal the divergence pattern of the Italian and Southern Alps lineages. Model comparison showed that the best fitting model was the SC model (AIC=1330.05), supporting the pattern of secondary contact between the Italian and Southern Alps lineages (Fig. 3a). In the second round of simulation, two modified versions of the best fitting basic model (SC), which include heterogeneous migration at different stages, were fitted as SC2M-

1 (gene flow concurrent with secondary contact) and SC2M-2 (gene flow later than secondary contact), for which the mean effect of selection at specific sites was quantified (Fig. 3b). The results showed that SC2M-2 was better supported than SC2M-1 (AIC_{SC2M-1}=1219.59, AIC_{SC2M-2}=1118.13). The likelihood ratio of SC2M-2 was also the lowest one among all comparisons in both simulation rounds (Fig. 3c and Table S2). Detailed parameters within SC2M-2 showed that the gene flow from Italian to Southern Alps (m_{SI}=2.96) and from Southern Alps to Italian (m_{IS}=2.33) were at the same level for most of the genome (average 96.8% of the loci, range 99.6%-90.9%). However, for a small part of the genome (average 3.2%, range 0.4%-9.1%) there was strongly asymmetric introgression from the Italian to the Southern Alps lineage (m_{eFI}=3.97), but effectively no asymmetric gene flow from the Southern Alps to the Italian lineage (m_{eIF}=5.76E-09). Genetic divergence was estimated to have accumulated during a period of isolation 5.9 times longer than the age of secondary contact (assuming a divergence time of 2.5 MYA between the two lineages (Salvi et al. 2013), this is estimated at~0.43MYA; Fig. 3d). Heterogeneous migration was estimated to have begun much later, approximately at 42.3% of the age of the first secondary contact.

Discussion

Here we have shown that introgression of a suite of sexually selected characters is accompanied by asymmetric introgression of a restricted part of the genome. This asymmetric introgression is more recent than the secondary contact between the two lineages, and the results suggest that the syndrome first spread throughout the Italian lineage from a geographic origin in the south of the lineage's contemporary distribution, before introgressing into the more distantly related lineage. These results demonstrate that sexual selection can cause a suite of correlated characters to spread throughout both closely and distantly related lineages with limited gene flow across the genome at large.

Sexual selection is a common cause underlying phenotypic differences between populations and species. Populations in Latium and Tuscany have been assigned to a distinct subspecies, *P. m. nigriventris*, based on expression of characters that we previously have shown to be under intra-sexual selection (While et al. 2015; Heathcote et al. 2016; MacGregor et al. 2017a). However, the genetic analyses revealed that populations with this phenotype are generally genetically more distant to each other than each of them is to neighbouring populations with the ancestral phenotype (i.e., *P. m. nigriventris* is an invalid taxonomic unit; also discussed in Bellati et al. 2011 and Salvi et al. 2013). This strongly suggests that the phenotypic characters have spread through introgressive hybridization. At first sight, sexual selection appears an unlikely driver of introgressive hybridization since divergence in sexual characters should promote reproductive isolation (Boughman 2001; Maan and Seehausen 2011). For example, female preference for males of the same lineage has been shown to restrict gene flow (e.g., Saetre et al. 1997; Boughman 2001; Seehausen et al. 2008; but see Stein & Uy 2006; Pfennig 2007; Lipshutz 2018; Tinghitella et al. 2018). However, when sexual selection mainly arises through male-male competition, there may be little that prevents hybridization upon secondary contact. Indeed, some of the best examples of selective introgression do in fact involve characters used in intra-sexual competition (e.g., Parsons et al. 1993; McDonald et al. 2001; Sluijs et al. 2008; Baldassarre et al. 2014; While et al. 2015; Lackey et al. 2018; Sluijs et al. 2018).

In common wall lizards, experimental data have demonstrated that differences in competitive ability between males with the sexually selected syndrome and males with the ancestral phenotype cause asymmetric hybridization in secondary contact (While et al. 2015). Since reproductive isolation is weak between the Italian and Southern Alps lineages (While et al. 2015), this competitive advantage should result in directional spread of phenotypes and genotypes from the Italian to the Southern Alps lineage. Our results, using simulations of

population demography and genomic cline analysis of more than 41,000 SNPs, show that the divergence in male competitive ability is indeed associated with asymmetric introgression, and that the morphological characters and colouration that are highly exaggerated in the Italian lineage have introgressed far beyond that expected based on neutral genetic markers (While et al. 2015; see also MacGregor et al. 2017b for an example of chemical composition of femoral secretions that does not introgress beyond that expected based on neutrality).

Assuming a divergence time since the beginning of the Pleistocene (Salvi et al. 2013), the reconstruction of the demographic history from these genomic markers suggests that the two lineages first exchanged genes around 400 thousand years ago. These estimates are uncertain, however, and the isolation and gene flow likely followed the contraction and expansion of species ranges during previous periods of climatic instability in this geographic region (e.g. Schönswetter et al., 2005; Sommer and Nadachowski 2006; Provan and Bennett 2008). Following secondary contact, most of the gene flow appears to have been occurring equally in both directions, which is consistent with the expectation that most genetic differences between even highly divergent lineages are selectively neutral. In contrast, a small part of the genome appears to have introgressed asymmetrically from the Italian lineage into the Southern Alps lineage. Importantly, there was no evidence for asymmetric gene flow in the opposite direction. These results are thus consistent with the experimental evidence for sexual selection and explain the mismatch between genotype and phenotype along the north-western coast of Italy (While et al. 2015).

Interestingly, the demographic simulations indicated that this directional gene flow is of more recent origin than the symmetrical gene flow. This suggests that the sexually selected syndrome was not present in the north of the Italian lineage at the time of secondary contact, but that it arrived later. The pairwise FST comparison and neighbour-joining tree analysis of genetic markers that associate with the syndrome indeed suggest that it might have originated

in the south of the distribution of the Italian lineage, and subsequently spread north within this lineage to eventually reach the hybrid zone between the Italian and Southern Alps lineages. Although this points towards a comparably recent origin of the syndrome, further data are necessary to establish more precisely its evolutionary origin and how its spread relates to the historical changes in climate, landmass, and population connectivity that have shaped the phylogeography of the species.

Adaptive introgression involves a repeated discarding of combinations of alleles and characters that do not work well together, keeping only those that are fit. A modular trait with a simple genetic basis should therefore be more likely to introgress under selection than would a suite of traits with more complex genetic architecture, since the latter may compromise developmental and functional integration in hybrids (Orr 1995). The phenotypic characters that comprise the sexually selected syndrome in wall lizards are continuous rather than discrete and likely to be influenced by a large number of loci. Indeed, hybrids in natural populations exhibit a continuum of colour phenotypes and morphologies. This makes the joint introgression of the suite of characters surprising. However, the characters are also highly correlated within populations of the Italian lineage, such that robust individuals with relatively large heads tend to have strong bite force, large testes, intense green dorsal colouration, extensive black ventral colouration, and lateral ocelli that reflect in the UV spectrum (MacGregor et al. 2017a). The suite of traits is also positively correlated with behavioural characters, both in staged interactions and in free-ranging lizards (Heathcote et al. 2016; MacGregor et al. 2017a). Since these characters tend to remain highly correlated in the hybrid zone (While et al. 2015), their development may have a shared genetic basis or involve genes that are physically linked. Our analyses identified a reasonably high number of candidate SNPs, representing about two percent of the genome, which showed consistent association with the syndrome phenotype. This number is consistent with the estimate of the

proportion of the genome that showed asymmetric introgression into the Southern Alps lineage (~3 %). However, the SNP markers produced by RAD-sequencing only provide a limited coverage of the genome, and most will not be causally linked to introgression of the phenotypes. More detailed studies are therefore necessary to identify candidate genomic regions and to explain the correlated introgression of suites of complex characters.

In summary, our results reveal that the suite of sexually selected characters of common wall lizards in western and central Italy originated within the Italian lineage and subsequently spread throughout this lineage before eventually reaching a hybrid zone with a highly divergent lineage. The continued introgressive spread of the sexually selected syndrome into this lineage is accompanied by asymmetric gene flow of about 3% of the genome, resulting in a mosaic genome in the receiving lineage.

Acknowledgements

We are grateful to all the students and volunteers who have contributed to this project through their assistance in the field (Sozos Michaelides, Hannah MacGregor, Nathalie Feiner, Natalia Zajac, Valentina Titone, Lindall Kidd, Ben Halliwell, Rachel Lewandowsky), Nathalie Feiner for discussion and three reviewers for very helpful comments on the manuscript. The computations were performed on resources provided by SNIC through the center for scientific and technical computing at Lund University under Project SNIC 2017/4-39. We thank Dr. Francois Bonhomme and Dr. Pierre-Alexandre Gagnaire for kindly providing the source code of the modified models for δaδi to detect heterogeneous gene flow, and Miguel Carneiro and Leif Andersson for granting access to the draft genome. This study was funded by the Swedish Research Council (E0446501). TU is supported by the Knut and Alice Wallenberg Foundations. DS is supported by the program 'Rita Levi Montalcini' for the recruitment of young researchers at the University of L'Aquila. Sample collection was

carried out in accordance with local laws and regulations in Italy, under the collection permit number Prot. PNM-2015-0009720.

References

Alexander D H, Novembre J, Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, **19**, 1655-1664.

Arnold M L. (2015). Divergence with genetic exchange. *Oxford University Press*.

Arnold M L, Kunte K. (2017). Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends in ecology & evolution*, **32**, 601-611.

Baldassarre D T, White TA, Karubian J et al.(2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution* **68**, 2644-2657.

Bellati A, Pellitteri-Rosa D, Sacchi R et al. (2011). Molecular survey of morphological subspecies reveals new mitochondrial lineages in *Podarcis muralis* (Squamata: Lacertidae) from the Tuscan Archipelago (Italy). *Journal of zoological systematic and evolutionary research*, **49**, 240-250.

Böhme W.(1986). Handbuch der Reptilien und Amphibien Europas, Band 2/II, Echsen (Sauria) III, Lacertidae III. Podarcis, 155-208.

Boughman J W. (2001). Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature*, **411**, 944-948.

Catchen J M, Amores A, Hohenlohe P et al. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, **1**, 171-182.

Chang C C, Chow C C, Tellier L Cet al. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.

Chen C, Liu Z, Pan Qet al.(2016). Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies Apis mellifera sinisxinyuan n. ssp. *Molecular biology and evolution*, **33**, 1337-1348.

Danecek P, Auton A, Abecasis G et al. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Derryberry E P, Derryberry G E, Maley J M et al. (2014). Hzar: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652-663.

Gruschwitz M, Böhme W. (1986). *Podarcis muralis* (Laurenti, 1786) – Mauereidechse. In: Böhme W (ed.) Handbuch der Reptilien und Amphibien Europas. Band 2/II. Echsen (Sauria) III (Lacertidae III: *Podarcis*). Wiesbaden, Aula, p 435.

Gutenkunst R N, Hernandez R D, Williamson S H et al. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Fan Z, Silva P, Gronau I et al. (2014). Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research*, **26**: 163-173.

Felsenstein J. (2005). PHYLIP: Phylogenetic inference program, version 3.6. *University of Washington, Seattle.*

Heathcote R J, While G M, MacGregor H E et al. (2016). Male behaviour drives assortative reproduction during the initial stage of secondary contact. *Journal of evolutionary biology*, **29**, 1003-1015.

Hedrick P W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular ecology*, **22**, 4606-4618.

Heliconius Genome Consortium. (2012). Genomic evidence for promiscuous exchange of adaptations among Heliconius butterfly species. *Nature*, **487**, 94-98.

Huang X, Kurata N, Wei X et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497-503..

Huerta-Sánchez E, Jin X, Bianba Z et al. (2014). Altitude adaptation in Tibetans caused by introgression ofDenisovan-like DNA. *Nature*, **512**, 194.

Izuno A, Kitayama K, Onoda Y et al.(2017). The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae). *Molecular ecology*, **26**, 1515-1532.

Kuijper B, Pen I, Weissing F J. (2012). A guide to sexual selection theory. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 287-311.

Lackey ACR, Martin MD, Tinghitella RM. (2018). Male competition and speciation: expanding our framework for speciation by sexual selection. *Current Zoology*, **64**, 69-73.

Lamichhaney S, Berglund J, Almén M S et al.(2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371.

Lipshutz S E. (2018). Interspecific competition, hybridization, and reproductive isolation in secondary contact: missing perspectives on males and females. *Current Zoology*, **64**, 75-88.

Maan M E, Seehausen O.(2011). Ecology, sexual selection and speciation. *Ecology letters*, **14**, 591-602.

MacGregor H E, While G M, Barrett J et al. (2017a). Experimental contact zones reveal causes and targets of sexual selection in hybridizing lizards. *Functional Ecology*, **31**, 742-752.

MacGregor H E, Lewandowsky R A, d'Ettorre P et al. (2017b). Chemical communication, sexual selection, and introgression in wall lizards. *Evolution*, **71**, 2327-2343.

MacGregor H E, While G M, Uller T. (2017c). Comparison of reproductive investment in native and non-native populations of common wall lizards reveals sex differences in adaptive potential. *Oikos*, **126**, 1564-1574.

McDonald D B, Clay R P, Brumfield R T et al. (2001). Sexual selection on plumage and behavior in an avian hybrid zone: experimental tests of male-male interactions. *Evolution*, **55**, 1443-1451.

Mitchell-Olds T, Willis J H, Goldstein D B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, **8**, 845-856.

Palmer D H, Kronforst M R. (2015). Divergence and gene flow among Darwin's finches: A genome‐wide view of adaptive radiation driven by interspecies allele sharing. *Bioessays*, **37**, 968-974.

Orr H A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, **139,** 1805-1813.

Parsons T J, Olson S L, Braun M J.(1993). Unidirectional spread of secondary sexual plumage traits across an avian hybrid zone (genus *Manacus*). *Science*, **260**, 1643-1646.

Peterson B K, Weber J N, Kay E H et al. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Pfennig K S.(2007). Facultative mate choice drives adaptive hybridization. *Science*, **318**, 965-967.

Provan J, Bennett K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in ecology & evolution*, **23**, 564-571.

Pardo-Diaz C, Salazar C, Baxter S W et al.(2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, **8**, e1002752.

Rougeux C, Bernatchez L, Gagnaire P-A. (2017). Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (*Coregonus clupeaformis*). *Genome biology and evolution*, **9**, 2057-2074.

Sætre G P, Král M, Bureš S.(1997). Differential species recognition abilities of males and females in a flycatcher hybrid zone. *Journal of Avian Biology*, 259-263.

Salvi D, HarrisJ, Kaliontzopoulou A et al. (2013). Persistence across Pleistocene ice ages in Mediterranean and extra-Mediterranean refugia: phylogeographic insights from the common wall lizard. *BMC Evolutionary Biology*, **13**, 1.

Seehausen O L E, Takimoto G, Roy D et al.(2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular ecology*, **17**, 30-44.

Sindaco R, Doria G, Razzetti E et al. (2006) Atlante degli anfibi e dei rettili d'Italia / Atlas of Italian amphibians and reptiles. Societas Herpetologica Italica, *Edizioni Polistampa*, Firenze.

Schönswetter P, Stehlik I, Holderegger R et al. (2005). Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, **14**, 3547-3555.

Schulte U, Veith M, Hochkirch A.(2012). Rapid genetic assimilation of native wall lizard populations (*Podarcis muralis*) through extensive hybridization with introduced lineages. *Molecular Ecology*, **21**, 4313-4326.

Sluijs I, Dooren T, Hofker KD et al. (2008) Female mating preference functions predict sexual selection against hybrids between sibling species of cichilid fish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2871-2877.

Sommer R, and Nadachowski A. (2006). Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review*, **36**, 251-265.

Song Y, Endepols S, Klemann N et al.(2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296-1301.

Stein A C, Uy J A C.(2006). Unidirectional introgression of a sexually selected trait across an avian hybrid zone: a role for female choice?*Evolution*, **60**, 1476-1485.

Tine M, Kuhl H, Gagnaire P-A et al. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications*, **5**, 5770.

Tinghitella RM, Lackey ACR, Martin M et al. (2018). On the role of male competition in speciation: a review and research agenda. *Behavioral Ecology*, **10**, 1093.

Van Der Sluijs I, Van Dooren T J, Hofker K D et al.(2008). Female mating preference functions predict sexual selection against hybrids between sibling species of cichlid fish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2871-2877.

Vilella AJ, Severin J, Ureta-Vidal A, et al. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware plylogenetic trees in vertebrates. *Genome Research,* **19**(2): 327-335.

Wellenreuther M, Svensson E I, Hansson B. (2014). Sexual selection and genetic colour polymorphisms in animals. *Molecular Ecology*, **23**, 5398-5414.

While G M, Michaelides S, Heathcote R J et al. (2015). Sexual selection drives asymmetric introgression in wall lizards. *Ecology Letters*, **18**, 1366-1375.

While G M, Uller T (2017). Female reproductive investment in response to male phenotype in wall lizards and its implications for introgression. *Biological Journal of the Linnean Society*, **121**, 876-882.

Wu T D, Reeder J, Lawrence M et al. (2016). GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Statistical Genomics: Methods and Protocols*, 283-334.

Data accessibility

Raw sequence reads from ddRAD-seq are deposited in the NCBI Sequence Read Archive (SRA) . The BioProject accession number is PRJNA486080.

Author contributions

TU and GMW conceived the study. All authors contributed to study design and data collection. WY conducted the laboratory work and the statistical analyses. WY and TU wrote the manuscript. All authors read, edited and approved the final submitted version of the manuscript.

Fig. 1 Summary of the results from the population genetic analyses. (a) A male *Podarcis muralis* from the Southern Alps lineage; (b) A male *Podarcis muralis* expressing the sexually selected syndrome; (c) Geographic locations of the populations included in this study. (d) Two-dimension PCA plots of all individuals and based on 22,446 unlinked SNPs; (e) Admixture clustering of sampled individuals into 2-5 groups. The proportion of each individual's genome assigned to each cluster is shown by the length of the coloured segments. (f) Neighbour-joining tree based on genetic distances measured by p-distance. Population abbreviations are as in panel C; (g) Maximum likelihood geographic clines of the

genetic hybrid index (HI, blue), candidate loci index for the sexually selected phenotype (CL, black), and the phenotypic indicator of the sexually selected phenotype (dorsal coloration or greenness, scale 1-10) for the populations across the hybrid zone from the Italian lineage (southernmost population is CN) to the Southern Alps lineage (westernmost population is LO). Dashed lines indicate the cline center for HI, CL, and greenness, respectively. Note that the genetic and phenotypic analysis were conducted on the same samples. Phenotypic clines for a larger set of male and female traits can be found in the Supplementary Information and in While et al. (2015).

Fig. 2 Detection and characterization of candidate loci associated with the sexually selected syndrome. (a) Distribution of nucleotide diversity ratios and F_{ST} values calculated for the RAD contigs. Green dots represent the candidate loci, and dashed lines indicate the top 5% of the values for nucleotide diversity and F_{ST} values; (b) Pair-wise F_{ST} among *P. muralis* population groups for candidate loci (green lines) and other ('neutral') loci (black lines). Green circles represent the Syndrome Group (S), and brown circles represent the Reference Group (R). Abbreviations indicate the location and Group membership of the populations as in Fig. 1C (SA-R: *LO,NL,SO*; SA-S: *ST,RA,GN*; TUN-R: *MV,MP*; TU: *CR,CN,CL*; TUE-R: *BL,BE*; LA-S: *SMA,RO,FU,FO*; LA-R: *DS,FG,QV,PS*); dashed lines indicate the center for hybrid index (HI) and greenness clines between Italian and Southern Alps lineages. (c) Neighbour-joining tree constructed by non-candidate ('neutral') loci for the populations in the Syndrome Group in Latium and Tuscany and their neighbouring populations. (d) Neighbour-joining tree constructed using the candidate loci for the same populations as in (c) Note that individuals with the sexually selected syndrome cluster together, the similarity to the results in Fig. 1F, and the fit with the geographic distribution of the populations. Numbers on edges are supporting rates for each major node from 1000 times of bootstrap.

Fig. 3 Demographic simulations of the historical gene flow between the Southern Alps and Italian lineages of *P. muralis*. (a) Four basic models were run for the first round of simulations: simple isolation (SI), isolation-with-migration (IM), ancient migration (AM), and secondary contact (SC). Black arrows represent gene flow. (b) Two updated models for the second round of simulations: heterogeneous migration concurrent with secondary contact (SC2M-1), and heterogeneous migration following secondary contact (SC2M-2). Black arrows represent the gene flow of neutral parts of the genome and green arrows represent the gene flow for parts of the genome under selection. (c) Akaike information criterion (AIC) for all the models in 20 runs (means ± SD). The best fitting model(s) are SC2M-2 (AIC =1118.13); (d) The best fitting model SC2M-2 shows significantly asymmetric introgression from the Italian lineage to the Southern Alps lineage.

Supporting information

Table S1 Sampling Information

Table S2 Phenotype data

Table S3 Hybrid Index for all populations

Table S4 Parameters for cline analysis

Table S5 Detailed parameters for demographic simulations

Fig. S1 Geographic clines for the greenness and blackness scores for males and females

Fig. S2 Geographic clines for hybrid index, greenness and individual candidate loci