SUPPLEMENTAL INFORMATION

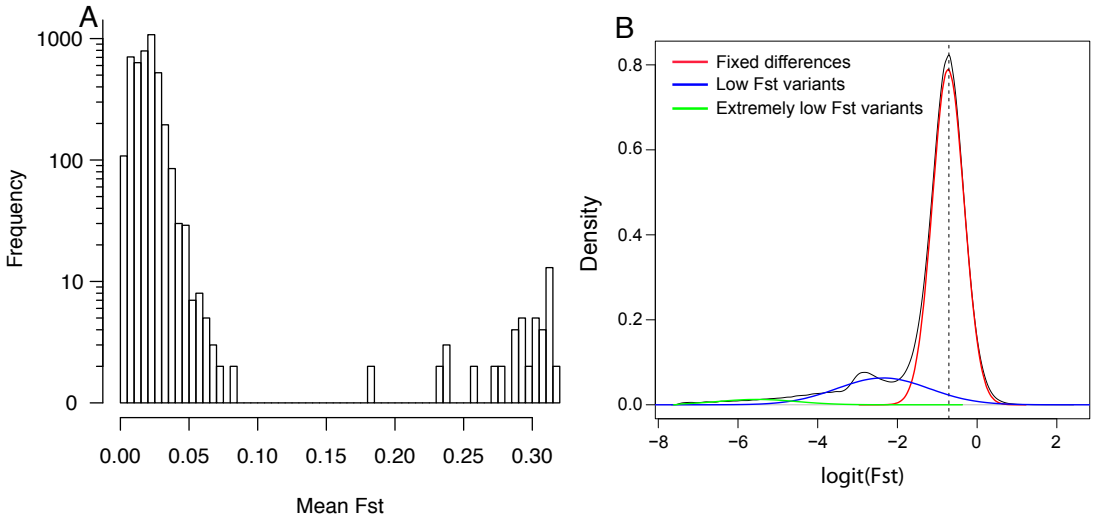

Figure S1, related to Figure 2 . A) Distribution of scaffold wide average F_{ST} among the ~6,000 assembled scaffolds. B) Distribution of variant F_{ST} values between white and tan birds within the *2 - 2^m* inversion. The distribution of F_{ST} is trimodal (χ^2 = 158, *df* = 3, p < 1e-50), with the largest component (red line; ~78.5% of variants) representing variants with F_{ST} ~ 0.33 (dashed vertical line). Other modes represent low F_{ST} variants (blue line, ~18.5% of variants), that are composed of polymorphisms segregating at low frequencies among either the white or tan birds, or extremely low frequency variants (green line, ~3% of variants) largely representing shared polymorphisms between white and tan birds.

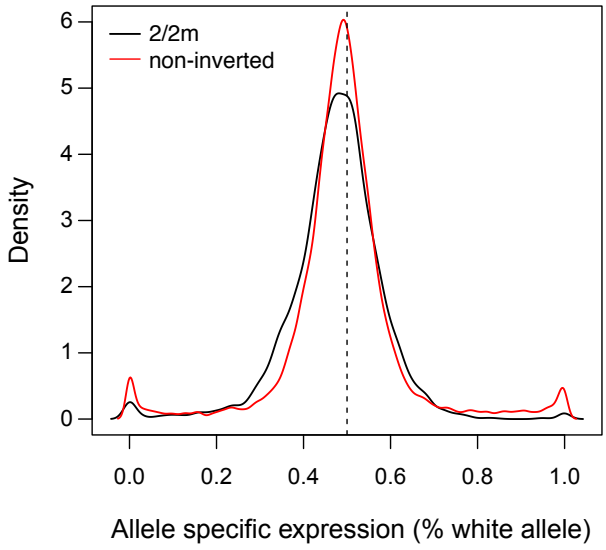

Figure S2, related to Figure 3. Distribution of allele specific expression values among three heterozygous white birds for variants inside (red line) and outside the *2/2^m* inversion. Dotted vertical line shows the expected allele specific expression value of 50%. In general, expression of *2^m*-linked genes is only mildly reduced relative to their *2*-linked homologues perhaps as a consequence of purifying selection acting to maintain proper dosage and the mitigating effects of double-recombination events or gene-conversion between *2* and *2^m*.

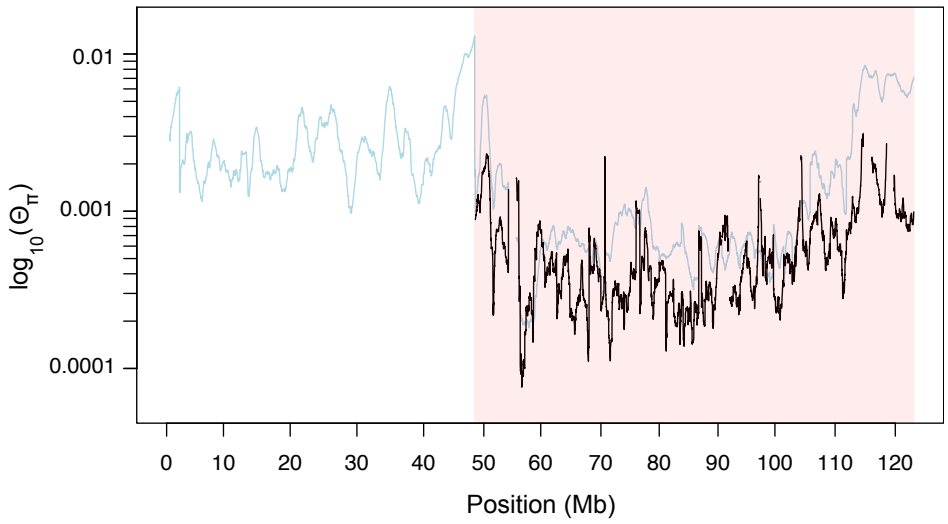

Figure S3, related to Figure 2. Sliding window estimates of diversity along chromosome 2^m (black line). For contrast, sliding window estimates of diversity along chromosome *2* are redrawn from Figure 2B (light blue).

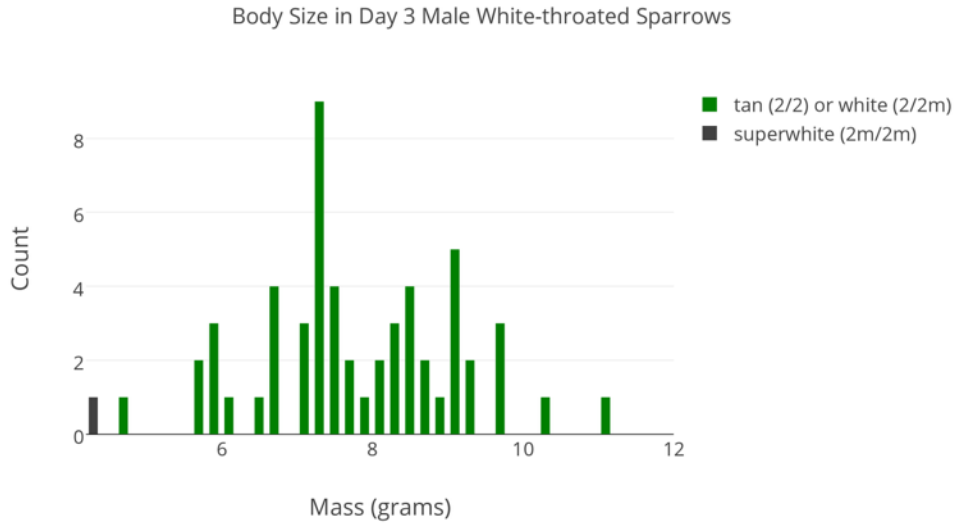

Figure S4, related to Figure 3. Body mass in 3 day old male chicks. One superwhite ($2^m/2^m$) male is the smallest male measured to date.

Morph	Tan (2/2)		White (2/2m)		Superwhite (2m/2m)		Total
Sex	male	female	male	female	male	female	
This Study	492	486	536	475	1*	2	1989
Thorneycroft[S1]	-	-	-	-	0	1	397
Michopoulus et al[S1]	-	-	-	-	0	0	58
Falls & Kopachena[S2]	-	-	-	-	1**	0	11
Horton et al.[S3]	-	-	-	-	0	1	602

Table S1, related to Figure 3. The frequencies of white-throated sparrows genotyped in this and other studies. Degradation of 2^m is caused by the rarity of homozygous ($2^m/2^m$) genotypes. The two $2^m/2^m$ females we identified were normal adult birds, as were the two females identified by Thorneycroft [S1] and Horton et al. [S3]. However, the one $2^m/2^m$ male chick we identified (*) was a runt from a WxW nest that was 45% below the average mass for a chick that age (>2 standard deviations below the mean), and 55% below the mass for its sibling (a tan male that was a day older), possibly suggesting that $2^m/2^m$ is lethal in males (see also Figure S3). Falls and Kopachena [S2] briefly report a $2^m/2^m$ male (**) but little information is given about its condition.

	WxW	WxT	TxW	TxT
Total # of Pairs	9	600	499	8
% of total pairs	0.8	53.8	44.7	0.7
% of disassortative pairs	-	54.6	45.4	-
% of assortative pairs	52.9			47.1
# of assortative pairs that are secondary	6			2
% of assortative that are secondary	67.0			25.0

Table S2, related to Figure 1. Pair type distribution for the Cranberry Lake population of white-throated sparrows for 1988 through 2014 (Male x Female). Polymorphism is maintained by disassortative mating. Only 1% of pairs are assortative, and of those, 47% are secondary, meaning that they form only after the initial pair member (who was of the opposite morph) disappeared.

Scaffold Name	Length	F_{st}	Order
NW_005081635.1	2,499,552	0.02	1
NW_005081812.1	685,102	0.02	1
NW_005081708.1	1,082,352	0.02	2
NW_005081569.1	6,428,133	0.31	2*
NW_005081577.1	5,965,185	0.31	3*
NW_005081742.1	823,114	0.33	4*
NW_005081832.1	313,837	0.31	5*
NW_005081548.1	11,897,188	0.33	6*
NW_005081699.1	1,154,867	0.31	7*
NW_005081620.1	3,174,558	0.33	8*
NW_005081574.1	5,686,916	0.33	9*
NW_005081589.1	4,450,299	0.33	10*
NW_005081596.1	4,337,631	0.33	11*
NW_005081611.1	3,579,899	0.33	12*
NW_005081596.1	4,337,631	0.33	13*
NW_005081591.1	4,472,293	0.33	14*
NW_005081642.1	2,137,341	0.33	15*
NW_005081561.1	7,563,313	0.33	16*
NW_005081632.1	2,519,554	0.33	17*
NW_005081596.1	4,337,631	0.33	18*
NW_005081561.1	7,563,313	0.33	19*
NW_005081582.1	4,953,235	0.33	20*
NW_005081569.1	6,428,133	0.31	21*
NW_005081654.1	1,929,533	0.30	22*
NW_005081553.1	10,164,019	0.26	23*

Table S3, related to Figure 2. Scaffolds and their order along chromosome 2 based on FISH mapping of BAC clones. Scaffold average F_{st} is provided for each. Scaffolds associated with BACs that mapped to the inverted region are highlighted with an asterisk.

	Pairtype (Male x Female)			
	WxW	WxT	TxW	TxT
Proportion of Male chicks ± SE (N)	0.32 ± 0.16 (6)	0.53 ± 0.02 (283)	0.48 ± 0.02 (241)	0.43 ± 0.16 (6)
Proportion of White chicks ± SE (N)	0.68 ± 0.14 (6)	0.49 ± 0.02 (283)	0.51 ± 0.02 (241)	0.26 ± 0.14 (6)
Proportion of EPP ± SE (N)	0.38 ± 0.18 (7)	0.28 ± 0.02 (199)	0.04 ± 0.02 (172)	0.41 ± 0.20 (6)

Table S4, related to Figure 3. Sex ratio, morph ratio and frequency of extra pair paternity (EPP) for the four different pair-types (N= number of pairs). White x White (WxW) assortative pairs produce female biased clutches and suffer a higher rate of EPPs. Tan x Tan pairs also suffer a higher rate of EPP relative to TxW pairs, which have the lowest rate of EPP. Note that white chicks found in TxT pairs are due to extra pair matings.

SUPPLEMETAL EXPERIMENTAL PROCEDURES

Field study: Field observations and samples were collected from a population of white-throated sparrows at the Cranberry Lake Biological Station in NY (44°07'N; 74°47'W) sampled between 1988 and 2014 (Master banding permit #22297 to E.M.T.). The field site supports ~100 pairs/season. Each pair was observed for a minimum of 80 hours over the entire season during which parental attendance at the nest, fledging, and predation were monitored. Blood samples were taken from nestling birds at post-hatch day 6, before they fledge the nest and while they are still completely dependent on parental provisioning. The sex of chicks was determined using standard protocols and primers [S5] and morph was determined using a PCR-based method, which amplifies the VIP gene and allows the identification of a morph-specific SNP [S2]. Paternity was determined using a combination of polymorphic microsatellite markers developed for other avian species [S6-S9]. To assign parentage we used the paternity analysis program CERVUS 3.0 [S10]. To test whether super white ($2^m/2^m$) birds are less common in the population that expected by chance we calculated their expected frequency based on the number of white male x white female (WxW) pairs and the frequency of extra-pair matings between white females and white males. Such extra pair matings should also produce 25% superwhite offspring. Based on our observations, WxW pairs comprise 0.8 % of the pairs in the population. Additionally, tan male x white (TxW) female pairs make up 44.7 % of the population but produce 4 % of offspring via extra pair mating with white males. Thus, based on sampling 1,989 birds from both within-pair and extra-pair matings, we would expect to see 12 superwhite chicks.

Fluorescent in situ hybridization: Candidate white-throated sparrow BAC clones (CHORI-264 [S11]) were identified using overgo probe hybridisation. BAC clones were grown and labeled following previously described protocols [S11,S12]. BAC DNA was labeled with SpectrumGreen®, SpectrumRed®, or SpectrumOrange® (Abbot Molecular, Inc. Abbot Park, Illinois) through nick translation. Homologous probes were hybridized to slides for 16 hours, heterologous probes and mixed heterologous/homologous probes were hybridized for 92 hours using probes of two different colours. Homologous probes were washed at 42 °C and all others were washed at 37 °C, following established stringency. Slides were

counterstained with DAPI III (Abbot Molecular Inc, Abbot Park, Illinois) and imaged using a multiphase microscope and fluorescent camera with CytoVision® software (Leica Biosystems, Germany).

Genome assembly: DNA was obtained from an adult tan morph male red blood cells (hematocrit) from one individual (karyotyped as 2/2, Z/Z). Another chromosomal polymorphism also segregates within the species, but is not correlated with morph [S1, S13, S14]. The sequenced individual was also homozygous for this polymorphism (3a/3a). The genome sequencing plan followed the recommendations provided in the ALLPATHS-LG assembler [S15]. This model requires 45x sequence coverage each of fragments (overlapping paired reads ~180bp length) and 3kb paired end reads as well as 5x coverage of 8kb PE reads. Total assembled sequence coverage of Illumina instrument reads was 63.3X using a genome size estimate of 1.09 Gb. This first draft assembly scaffold gaps were closed where possible with mapping of the same reference assembly Illumina sequences and local gap assembly. Contaminating contigs, trimmed vector in the form of X's and ambiguous bases as N's in the sequence were removed. All contigs 200bp and smaller were removed.

Gene annotation: Prior to gene annotation the assembly was first masked with known repeats using RepeatMasker [S16] and WindowMasker [S17]. Annotation was done using the NCBI gene annotation pipeline which is fully described and publicly available (http://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/). The essential modules of this pipeline are the alignment programs Splign ([S18]; cDNA) and ProSplign (protein) and an HMM-based gene prediction program Gnomon (http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.shtml). Following manual checks of the gene annotation output we predict a total sparrow gene count of 15,121 with a mean length of 595 amino acids, similar to other avian genomes.

Genome resequencing: Twenty-four tan and 25 white morph males sampled above were chosen for resequencing. DNA was extracted using Qiagen DNeasy kits and samples were pooled in equimolar amounts for library generation. Library preparation for each pool was done using Illumina TruSeq Sample Prep Kit, and sequencing was done using TruSeq SBS kit version 3 and data was processed with Cassava 1.8.2. Sequencing was done on two lanes of a HiSeq2000, yielding 726,657,908 100bp paired end reads. Three additional white morph individuals were individually sequenced. These three individuals were previously collected in Chicago, IL, US [S19]. Library preps were done using the Kapa Biosystems Library Construction Kit and sequenced using their TruSeq Rapid SBS sequencing kit version 1. Samples were sequenced on a single lane of a HiSeq2500 and yielded over 65 million reads per sample and a total of 416,496,508 paired end (100bp) reads. We also sequenced whole genome of two outgroups, Harris' sparrow (*Zonotrichia querula*) and golden-crowned sparrow (*Zonotrichia atricapilla*) using specimens housed at the University of Alaska Museum. Library preparation for these species was done using the Kapa Biosystems Hyper Kit and TruSeq Rapid SBS sequencing kit version 2. Sequencing was then conducted on an Illumina Hi-Seq 2500 and yielded over 240 million paired-end (120 bp) reads per species.

Phylogenomic analysis: To place chromosomes *2* and *2^m* into a phylogenetic context, we compared white-throated sparrow genomic data with mRNA-seq data from rufous-collared sparrow *(Zonotrichia capensis)*, white-crowned sparrow (*Zonotrichia leucophrys*) [S19], song sparrow (*Melospiza melodia*) [S19] and dark-eyed junco (*Junco hyemalis*) [S20] and whole genome sequencing of Harris' sparrow (*Zonotricia querula*) and golden-crowned sparrow (*Zonotricia atricapilla*). To do so, we mapped mRNA and gDNA reads from each outgroup species to the tan reference genome using *bwa mem* [S21]. gDNA reads from the tan pool of birds was mapped to the tan reference genome as described above; gDNA reads from the white pool of birds was mapped to the composite reference genome as described below (see '*Population genomic analysis'*). We then generated a super matrix of any informative sites and encoded sites that are polymorphic within species with their corresponding IUPAC ambiguity code. We partitioned this super matrix into two regions that either contain the *2/2^m* inversion region or do not. The phylogeny of these species was inferred for each partitioned super matrix using RAxML [S22] with the GTRCAT mutation model. Bootstrap support values for each node were generated with 1000 bootstrap replicates. The topology produced above was compared against a tree where the tan and white morphs were

constrained as monophyletic. The Tan+White tree was built in RAxML [S22] under a CAT model with 100 random addition sequence (RAS) replicates, and support values were obtained via 100 bootstrap replicates. The likelihood of each tree given our data was obtained using FASTTREE 2 [S23], and following comparison methods were employed in CONSEL [S24]: approximately unbiased test (AU), two bootstrapping methods (NP & BP), Bayesian posterior probabilities (PP), Kishino-Hasegawa (KH), Shimodaira- Hasegawa (SH), weighted Kishino-Hasegawa (WSH), and weighted Shimodaira-Hasegawa (WSH). P-values were calculated for each metric and used to assess which topology best describes the data and whether the trees are significantly different

Population genomic analysis: Pooled reads were mapped to the assembled genome using *bwa mem* [S21] and genome-wide estimates of variant allele frequencies were estimated with *CRISP* [S25] or *mpileup* [S26]. Mapping and variant detection was performed with default base and mapping quality thresholds and at least two reads with alternative alleles were required to call a site as polymorphic. On average, read depth in both pooled samples was ~27.5X, and sites with low read depth (<5X) or high read depth (>80X) were discarded. Variants were annotated using *SnpEffect* [S27]. F_{ST} was estimated between tan and white pooled samples as,

$$F_{ST} = \frac{H_{tot} - H_{with}}{H_{tot}}$$

where H_{tot} is average heterozygosity between tan and white birds and H_{with} is average heterozygosity within tan and white birds. Sliding window estimates of F_{ST} were calculated in 100Kb windows with a 10Kb step size. Windows with less than 100 variants were excluded from analysis.

To estimate the fraction of variants within the inversion region that have F_{ST} ~0.33 (the expected value for variants that are fixed between tan and white birds), we modelled $logit(F_{ST})$ as a mixture of three normal distributions using the *mixtools* [S28] package in R [S29]. To avoid spurious parameter estimates, we removed variants with F_{ST} less than 0.0005. Parameters of the mixture distribution (mixture parameter, mean, and standard deviation) were estimated by randomly sampling 100 sets of 10,000 variants within the inversion and calculating the average estimate across these random sets of SNPs.

Pairwise genetic diversity (Θ_π) was estimated using *PoPoolation* [S30]. To estimate Θ_π on chromosome *2*, we utilized allele frequency estimates from the tan pool of birds. To estimate Θ_π on chromosome 2^m, we generate a pseudo-white reference genome by substituting any mutation at ~50% frequency within the $2/2^m$ region in the pool of white birds into the tan reference genome using GATK's FastaAlternativeReferenceMaker [S31]. We then performed competitive remapping by mapping reads from the white pool of birds to a composite reference genome containing both the reference tan and pseudo-reference white genomes. Competitive remapping of the white-pool successfully separated reads emanating from either chromosome *2* or 2^m (median read depth of *2* scaffolds = ~12X; median read depth of 2^m scaffolds = ~12X). As a negative control, we also performed competitive mapping of reads from the tan pool of birds to the composite genome (median read depth of *2* scaffolds = ~27X; median read depth of 2^m scaffolds = 0X). Sliding window estimates of Θ_π were calculated in sliding 1Mb windows with a 10Kb step size. Windows with less than 100 variants were excluded from analysis.

We estimated pair-wise linkage disequilibrium (r^2) between pairs of SNPs within the sequence reads using the 'direct estimate' method of LDx [S32]. We only considered sites with intersecting read depth greater than 5 and minor allele frequency at either locus > 15%. These thresholds have been found to produce accurate estimates of r^2 [S32]. We calculated average r^2 in 1bp distance bins separately for regions of the genome outside the inversion and inside the inversion for 2 and 2^m.

We calculated the Direction of Selection (DoS) statistic [S33] for each gene in the genome using either Harris' or golden-collared sparrow as outgroup. To calculate *dn* and *ds* for each outgroup, we mapped gDNA reads to either the tan reference genome (as above) or the white pseudo-reference genome. To

calculate *pn* and *ps* for either *2*, 2^m, or the remainder of the genome we only considered polymorphisms with minor allele frequency less than 15% with at least four reads supporting the polymorphism. The use of low frequency polymorphisms was necessary to ensure that polymorphisms contributing to *pn* and *ps* were not fixed differences between *2* and 2^m. To calculate DoS between *2* and 2^m we treated fixed differences between them as *pn* and *ps*. To estimate average DoS for various classes of genes (see Figure 2A) we calculated the weighted average DoS and weighted standard error [S33]. We calculated Fay and Wu's *H* in non-overlapping 50Kb windows genome-wide using either Harris' or golden-collared sparrow as outgroup species to polarize the allele frequency spectrum from the tan pool. Fay and Wu's *H* was estimated using *npstat* [S34].

Gene expression analysis. We examined patterns of gene expression in white and tan birds for genes inside and outside the inverted region. We utilized previously published *mRNA*-seq data from whole brain extracts from three white- and three tan-morph birds [S19]. We first mapped the *mRNA*-seq from each bird to the tan reference genome using *bwa mem*. Next, we constructed masked reference genome for each of the six birds based on polymorphisms identified in this first round of mRNA-seq mapping using *bedtools maskfasta* [S35] function. We then remapped *mRNA*-seq to these individually masked genomes using tophat2 [S36] and estimated differential expression using cuffdiff [S37].

Allele specific expression analysis. We examined patterns of allele specific expression (ASE) within and outside the inversion for three heterozygous ($2/2^m$) white birds. We utilized previously published *mRNA*-seq data from whole brain extracts [S19] and also resequenced genomic DNA from the same three birds. To avoid reference allele bias in ASE estimation, we first mapped *gDNA* of these three birds to the tan reference genome. We then masked (i.e., converted the reference allele to 'N') all sites in the reference genome that were found to be heterozygous in any of the three birds using *bedtools maskfasta* [S35] function. We mapped *mRNA*-seq data to the masked reference genome using *bwa mem* [S21] and estimated ASE as the fraction of 'tan' alleles (e.g., the reference allele in the unmasked genome) at exonic SNPs with minimum *mRNA*-seq read depth of 50 and maximum *gDNA*-seq read depth of 15. We further required that SNPs used for ASE estimation were heterozygous in all three birds. ASE estimates per SNP were then averaged across the three birds.

Confirmation of pooled genotyping. To confirm fixed difference between morphs identified in pooled genotypes, we resequenced the genomes of three individual white birds sampled from a different population than the pooled sample. As expected, 75% variants within the inversion are heterozygous in all three birds; note, given limited sequencing depth of individual birds (~10X) we expect that a fraction of truly polymorphic variants to appear homozygous thus explaining the discrepancy between the rate of fixed differences (75% versus 78.6%) observed in our individual and pooled based resequencing, respectively.

SUPPLEMENTAL REFERENCES

S1. Thorneycroft H.B. (1975). Cytogenetic study of white-throated sparrow, *Zonotrichia-albicollis* (Gmellin). Evolution. *29*, 611-621.

S2. Michopoulos V., Maney D.L., Morehouse C.B., and Thomas J.W. (2007). A genotyping assay to determine plumage morph in the white-throated Sparrow (*Zonotrichia albicollis*). Auk. *124*, 1330-1335.

S3. Horton, B. M., et al., Behavioral Characterization of a White-Throated Sparrow Homozygous for the ZAL2(m) Chromosomal Rearrangement. *Behav Genet* **43**, 60-70 (2013)

S4. Falls, J. B. Kopachena, J. G (2010) in *The birds of North America Online* (ed Poole, A.) (Cornell Laboratory of Ornithology, Ithaca).

S5. Griffiths R., Double M.C., Orr K., and Dawson R.J.G. (1998). A DNA test to sex most birds. Mol Ecol. *7*, 1071-1075.

S6. Poesel A., Gibbs H.L., and Nelson D.A. (2009). Twenty-one novel microsatellite DNA loci isolated from the Puget Sound white-crowned sparrow, Zonotrichia leucophrys pugetensis. Molecular Ecology Resources. *9*, 795-798.
S7. Dawson R.J.G., Gibbs H.L., Hobson K.A., and Yezerinac S.M. (1997). Isolation of microsatellite DNA markers from a passerine bird, Dendroica petechia (the yellow warbler), and their use in population studies. Heredity. *79*, 506-514.
S8. Petren K. (1998). Microsatellite primers from Geospiza fortis and cross-species amplification in Darwin's finches. Mol Ecol. *7*, 1782-1784.
S9. Jeffery K.J., Keller L.F., Arcese P., and Bruford M.W. (2001). The development of microsatellite loci in the song sparrow, Melospiza melodia (Aves) and genotyping errors associated with good quality DNA. Mol Ecol Notes. *1*, 11-13.
S10. Kalinowski S.T., Taper M.L., and Marshall T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. *16*, 1099-1106.
S11. Lear T.L., Lundquist J., Zent W.W., Fishback J., W.D., and Clark A. (2008). Three autosomal chromosome translocations associated with repeated early embryonic loss (REEL) in the domestic horse (Equus caballus). Cytogenetic and Genome Research. *120*, 117-122.
S12. Lear T.L., Brandon R., Piumi F., Terry R.R., Gu√©rin G., Thomas S., and Bailey E. (2001). Horse gene mapping report: Mapping of 31 horse genes in BACs by FISH. Chromosome Res. *9*, 261-262.
S13. Thorneycroft H.B. (1966). Chromosomal polymorphism in white-throated sparrow Zonotrichia albicollis (Gmelin). Science. *154*, 1571-157
S14. Romanov M., N, Dodgson J., B, Gonser R., A, and Tuttle E., M. (2011). Comparative BAC-based mapping in the white-throated sparrow, a novel behavioral genomics model, using interspecies overgo hybridization. BMC research notes. *4*, 211-211.
S15. Gnerre S., MacCallum I., Przybylski D., Ribeiro F.J., Burton J.N., Walker B.J., Sharpe T., Hall G., Shea T.P., Sykes S. et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA. *108*, 1513-1518.
S16. Smit A.F.A., Hubley R., and Green P. (1996). RepeatMasker Open-3.0.
S17. Morgulis A., Gertz E.M., Schaeffer A.A., and Agarwala R. (2006). WindowMasker: Window-based masker for sequenced genomes. Bioinformatics. *22*, 134-141.
S18. Kapustin Y., Souvorov A., Tatusova T., and Lipman D. (2008). Splign: Algorithms for computing spliced alignments with identification of paralogs. Biology Direct. *3*, 20.
S19. Balakrishnan C.N., Mukai M., Gonser R.A., Wingfield J.C., London S.E., Tuttle E.M., and Clayton D.F. (2014). Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. PeerJ. *2*, e396.
S20. Stager M., Swanson D.L., and Cheviron Z.A. (2015). Regulatory mechanisms of metabolic flexibility in the dark-eyed junco (Junco hyemalis). J Exp Biol. *218*, 767-777.
S21. Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997
S22. Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. *30*, 1312-1313.
S23. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.
S24. Shimodaira, H., and Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247
S25. Bansal V. (2010). A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. *26*, i318-i324.
S26. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., and Durbin R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics. *25*, 2078-2079.
S27. Cingolani P., Platts A., Wang L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., and Ruden D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. Fly. *6*, 80-92.

S28. Benaglia, T., Chauveau, D. R. Hunter, D. R. Young, D. S. Mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software* **32**, 1-29 (2009).
S29. R Core Development Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, 2014).
S30. Kofler R., Orozco-terWengel P., de M., N., Pandey R.V., Nolte V., Futschik A., Kosiol C., and Schlotterer C. (2011). Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE. *6*, e15925.
S31. McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M. et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297-1303.
S32. Feder A.F., Petrov D.A., and Bergland A.O. (2012). LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. PLoS ONE. 7.
S33. Stoletzki N., and Eyre-Walker A. (2011). Estimation of the neutrality index. Mol Biol Evol. 28, 63-70.
S34. Ferretti, L., Ramos-Onsins, S.E., Perez-Enciso, M. (2013). Population genomics from pool sequencing. Mol Ecol *22,* 5561-5576.
S35. Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
S36. Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., and Salzberg S.L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14.
S37. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L., and Pachter L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols. 7, 562-578.