Genomics of introgression in the Chinese horseshoe bat (*Rhinolophus sinicus*) revealed by transcriptome sequencing

Xiuguang Mao[1,2,*], Georgia Tsagkogeorga[2], Sebastian E. Bailey[2], Stephen J. Rossiter[2,*]

1 Institute of Estuarine and Coastal Research, East China Normal University, Shanghai 200062, China
2 School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

*Corresponding author: xgmao@sklec.ecnu.edu.cn; s.j.rossiter@qmul.ac.uk

Subtitle: Genomics of introgression in horsesoe bats

published as Mao et al. (2017) Biological Journal of the Linnean Society, 121, 698–710,

ABSTRACT

Recent genomic studies show that introgression can occur at a genome-wide scale among recently diverged lineages. However, introgression is difficult to distinguish from incomplete lineage sorting (ILS), and these processes are expected to occur together. Moreover, ncDNA introgression is less easily detected than mtDNA introgression, and as such its prevalence is less well understood. The Chinese horseshoe bat (*Rhinolophus sinicus*) occurs as three distinct forms on mainland China: the subspecies *R. s. septentrionalis* and two parapatric clades of *R. s. sinicus* (Central and East *R. s. sinicus*). Previous work suggested widespread mtDNA introgression between these subspecies, however, no ncDNA introgression was detected. In this new study we sampled the coding genomes of all three forms of *R. sinicus* in order to perform a more sensitive test for ncDNA introgression against an expected background of incomplete lineage sorting. We assembled 3548 nuclear protein-coding genes from these and three congeneric species, and built a high confidence species tree using Maximum Likelihood and Bayesian concordance methods. Phylogenetic analysis suggested a mosaic genome for Central *R. s. sinicus* derived from *R. s. septentrionalis* and East *R. s. sinicus*. Nuclear DNA introgression between Central *R. s. sinicus* and *R. s. septentrionalis* was supported by three different tests, whereas ILS could not be ruled out completely. Our findings, in line with other recent results, indicate that recently diverged taxa undergo large scale secondary introgression, and that this process likely operates alongside incomplete lineage sorting to give rise to phylogenomic discordances or even mosaic genomes.

Additional Keywords: RNA-Seq - speciation - introgressive hybridization - cytonuclear discordance - *Rhinolophus*

INTRODUCTION

Hybridization is very common between diverging populations (reviewed in Abbott *et al*., 2013). Increasing numbers of studies have revealed the role of hybridization in driving functional novelty and adaptive introgression (see Seehausen, 2004; Song *et al*., 2011; Clarkson *et al*., 2014; Norrisa *et al*., 2015), as well as in speciation (see Mallet, 2008; Mavárez & Linares, 2008; Nolte & Tautz, 2010; Brandvain *et al*., 2014; reviewed in Abbott *et al*., 2013;). Studies on genomics of speciation have benefited greatly from the advent of high throughput sequencing, which has allowed for genome-wide markers to be rapidly and cheaply generated from non-model organisms (Shendure & Ji, 2008; Hudson, 2008; Hawkins *et al*., 2010; Payseur & Rieseberg, 2016). Transcriptome sequencing (RNA-seq), in particular, has been widely used for obtaining protein coding gene sequences for evolutionary analysis of neutral and adaptive variation (Alvarez *et al*., 2015; Tepolt & Palumbi, 2015), with specific applications ranging from scans for genes and sites undergoing accelerated differentiation or positive selection (e.g. Phifer-Rixey *et al*., 2014; Marra *et al*., 2014; Westram *et al*., 2014; Pereira *et al*., 2016) to tests for topological conflicts in phylogenomic analysis (e.g. Zhang *et al*., 2013; Fernández *et al*., 2014; Sharma *et al*., 2014) as well as for hybridization and introgression among diverged taxa (e.g. Cui *et al*., 2013; Zhang *et al*., 2013; Hodgins *et al*., 2014; Stuglik & Babik, 2016).

Closely related taxa offer good opportunities for studying the process of reproductive isolation in the early stages of divergence, when hybridization and introgression still occur among taxa. However, it is often difficult to distinguish introgression from incomplete lineage sorting (ILS) (Clark, 1997), which is also expected to occur in recently diverged taxa (e.g. Liu *et al*. 2014; Marcussen *et al*. 2014). One approach for detecting the presence of introgression against a background of ILS is by screening for topological variation among gene genealogies. Introgressive hybridisation will lead to greater numbers of shared derived alleles among populations exchanging genes than would expected under a scenario of ILS acting alone (Durand *et al*. 2011). Additional information can be gained by examining the genomic distribution of sets of loci supporting particular topologies (see Martinsen *et al*., 2001; Schumer *et al*., 2013) based on the rationale that loci that have undergone recent hybridization will tend to be clustered in the genome (Martinsen *et al*., 2001). However, this usefulness of this approach is limited because recombination quickly erodes linkage disequilibrium (see Schumer *et al*., 2013).

Bats account for a fifth of all mammal species, yet reported cases of introgression in bats were , until very recently, rare (e.g. Webb & Tidemann, 1995; Hoffmann *et al*., 2003; Berthier *et al*., 2006; Larsen *et al*., 2010; Neubaum *et al*., 2007; Mao *et al*., 2010, 2013a, 2013c; Bogdanowicz *et al*., 2009, 2012; Vallo *et al*., 2012). However, with one exception (Chattopadhyay *et al*., 2016) these previous studies typically relied on one or only on a small number of loci and were thus unable to confidently disentangle genetic signatures of putative introgression from those of ILS. Similarly, as in other vertebrate groups (e.g. birds, Pons *et al*., 2014), small numbers of nuclear markers are likely to have also led to relatively fewer reported cases of ncDNA than of mtDNA in bats.

To investigate whether widespread nuclear introgression can occur in the history of divergence of closely related taxa, we studied the Chinese horseshoe bat (*Rhinolophus*

sinicus) based on a genome-wide scale dataset. This species occurs as three distinct forms on mainland China: the subspecies *R. s. septentrionalis* and two geographical clades of *R. s. sinicus* (Central and East *R. s. sinicus*) (see Figure 1a, also see Mao *et al*., 2013b). *R. s. septentrionalis* is significantly larger and genetically divergent from *R. s. sinicus* (see Csorba *et al*., 2003; Mao *et al*., 2013c) and, similarly, the two forms of *R. s. sinicus* can be separated on the basis of skull morphology (Xu *et al*., 2005) as well as both mtDNA and nuclear DNA sequences, although these have not yet been formally described as separate taxa (see Mao *et al*., 2013b and this study). Phylogenetic analyses of this species complex using approximate Bayesian computation (ABC) applied to two mitochondrial genes, four nuclear and seven microsatellite loci, supported a scenario in which *R. s. septentrionalis* diverged first from *R. s. sinicus* lineages followed by the subsequent divergence of Central and East *R. s. sinicus* (Mao *et al*., 2013b). The overlapping ranges of these two subspecies are therefore best explained by secondary contact. Analyses of gene flow and phylogenetic topology also indicated widespread mtDNA introgression from *R. s. septentrionalis* into Central *R. s. sinicus*, and, although no such ncDNA introgression was detected, it cannot be ruled out completely given the limited power of detection of the small number of loci tested (Mao *et al*., 2013b).

The main aim of this study was to assess the extent of nuclear introgression among the three focal lineages. For this, we generated RNA-seq data for three focal taxa (*R. s. septentrionalis*, Central *R. s. sinicus* and East *R. s. sinicus*) and three other horseshoe bats (*R. pearsoni*, *R. rex* and *R. ferrumequinum*). We assembled alignments for 3,548 nuclear protein-coding genes, to which we also added the published orthologous sequences from the echolocating bat *Myotis lucifugus* and the non-echolocating fruit bat *Pteropus vampyrus* as outgroups. Based on these genome-wide datasets, we first built a high confidence phylogenomic tree using both concatenated and Bayesian concordance methods. We then used three different methods to test for contrasting histories among gene genealogies, which can be used to signify the occurrence of introgression against a background of ILS. Finally, we assessed the relative contributions of ILS and introgression on genome divergence by examining the genome distribution patterns of loci supporting different tree topologies reconstructed for the three focal taxa.

MATERIAL AND METHODS

SAMPLING

We sampled one individual each from three Chinese mainland forms of *R. sinicus* (Central *R. s. sinicus*, East *R. s. sinicus* and *R. s. septentrionalis*, see Figure 1a) and three congeneric species (*R. pearsoni*, *R. rex* and *R. ferrumequinum*). Individuals of the three *R. sinicus* were sampled from parapatric regions (see Figure 1a). Each individual was euthanized and sampled for muscle, brain and heart tissue. Bats sampling was approved by the Animal Ethics Committee of East China Normal University (approval ID AR201003002) and the local Protection and Research. Tissue samples were flash frozen in liquid nitrogen in the field, and transferred to a -80 °C freezer until RNA extraction.

RNA EXTRACTION, SEQUENCING AND SEQUENCE ASSEMBLY

For each specimen, total RNA was extracted from the mix of three tissue types. All steps of RNA extraction, mRNA purification, cDNA library construction (insert fragment of ~250 bp) and sequencing using an Illumina HiSeq 2000 sequencer (100 bp paired-end) were performed by BGI. Raw reads of RNA-seq were cleaned by removing adaptors as well as low quality bases (Phred quality score < 20) using NGSQCToolkit_v2.3.3 (Dai *et al.,* 2010). The quality of the paired-end sequences was controlled before and after the trimming process using FastQC (version 0.10.1, www. bioinformatics.babraham.ac.uk/projects/fastqc/). The data sets supporting the results of this article are available at the Sequence Read Archive repository (SRA accession numbers: SRR5219050, SRR5219064, SRR5219072-73, SRR5219075-76).

For each specimen, we performed d*e novo* transcriptome assembly using *Trinity* version 2011-08-20 (Grabherr *et al*., 2011) under default parameters. The abundance of the assembled contigs was estimated by mapping reads back to the contigs, using RSEM and the Bowtie aligner, implemented in the *Trinity* software package. Contigs with average per-site coverage of <10x and sequence length of < 200 bp were excluded from further analysis.

IDENTIFICATION OF ORTHOLOGOUS GENES

To identify orthologous genes across our assembled transcriptomes, we used a two-step approach. First, we used Ensembl BioMart to obtain all one-to-one orthologous genes between human and two bat species, *Pteropus vampyrus* and *Myotis lucifugus*. For each gene, only the longest Coding DNA Sequence (CDS) was retained, resulting in a total of 11,158 one-to-one orthologous CDSs in human. These collection of CDSs were then used as queries in similarity searches against the transcriptome assemblies for each individual, using TBLASTX under an E-value < 10^{-6} and a nucleotide sequence identity cut-off of 75%. The longest isoforms of the best hit were retrieved as 'genes', and only genes shared by all six horseshoe bats were retained for further analysis. In addition to these horseshoe bats, we also included the nuclear sequence data from *Myotis lucifugus* and *Pteropus vampyrus*. Scripts used in generating genes shared by all six horseshoe bats are available in a Dryad Repository (xxxxxx).

Multiple sequence alignments (MSAs) based on nucleotides were performed using MAFFT version 5 (Katoh *et al*., 2007) with default parameters. Alignments were trimmed for gaps at both 5´and 3´ends, and checked for the presence of premature stop codons by translating into amino acid sequences. Only genes with valid Open Reading Frames (ORFs) were retained. Based on these sequence alignments, we further checked for the presence of potential paralogous genes by measuring heterogeneity (i.e. the mean sitewise Shannon entropy) for each gene sequence alignment using Shannon Heterogeneity In Alignments tool Version 1.1 (available on request from J. Parker). Sequence alignments with mean sitewise Shannon entropy values of > 0.02 were checked manually for potential paralogous copies and removed. We also excluded sequence alignments with mean sitewise Shannon entropy values of < 0.001 because there were almost no variable sites in these alignments.

Mitochondrial genes for each *Rhinolophus* taxon were identified by performing BLASTN (the E-value < 10^{-6}) using assembled contigs for each taxon as queries and the mitochondrial genome of *R. ferrumequinum* as the reference database (NCBI Reference Sequence: NC_016191.1). Mitochondrial gene sequences from *Pipistrellus abramus* (GenBank accession: AB061528.1) and *Rousettus aegyptiacus* (GenBank accession: AB205183.1) were included as outgroups.

INFERENCE OF SPECIES TREE

To infer the species tree we used two methods. First we generated a maximum-likelihood (ML) tree based on nucleotide sequences from 3,548 concatenated nuclear protein-coding genes in the software RAxML 7.2.8 (Berger *et al*., 2011). For this analysis, we used the GTRGAMMA model and bootstrap support was estimated from 1000 pseudoreplicate searches. Second we performed a Bayesian concordance analysis (see Ané *et al*., 2007) using the software BUCKy (Larget *et al*., 2010). This method addresses some shortcomings of standard concatenation methods, which recent studies indicate can lead to erroneous support for incorrect hypotheses (e.g. Kubatko & Degnan, 2007; Weisrock *et al*., 2012; Nosenko *et al*., 2013; Salichos & Rokas 2013). Such spurious phylogenetic inferences arise from discordance among gene trees due to processes such as incomplete lineage sorting (ILS) and/or introgressive hybridization. Bayesian concordance analysis does not assume anything about the underlying causes of gene tree conflicts, and works by clustering groups of compatible gene trees in order to identify the dominant signal in the data (Ané *et al*., 2007). To implement BUCKy, we first used MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) to obtain the posterior probability of each gene tree for those loci present in all eight taxa (n = 2,919). We performed two Metropolis-coupled Markov chain Monte Carlo (MCMC) runs, each comprising four chains and 10 million generations. Trees were sampled every 100 generations, and the first 25% of the sampled trees was discarded as a burn-in. BUCKy was run to combine all genes and estimate the concordance factor (CF) for each group under different values of the Dirichlet parameter (α = 1, 2 and 5). Output files from results were visualized with bucky-tools (www. stat.wisc.edu/~ane/bucky/).

mtDNA PHYLOGENY

To assess relationships among the focal taxa based on mitochondrial haplotypes, we assembled and aligned sequence data from concatenated 13 mitochondrial protein-

coding genes. We undertook ML-based phylogenetic tree reconstruction using RAxML with the same parameters as the above. By comparing the results of the mtDNA tree and species tree, we assessed the evidence for mito-nuclear discordance, as previously suggested from analyses of two mitochondrial gene sequences (Mao *et al*., 2013b).

TESTS FOR INTROGRESSION

To test for evidence of introgression among the three focal taxa, we applied three methods. First, we used the ABBA-BABA test based on the calculation of Patterson's D-statistic, which is an explicit test of gene flow versus ILS (Green *et al*., 2010; Durand *et al*., 2011). Given a four-species phylogeny (((P1,P2),P3),O) in which P1, P2 and P3 are defined as three closely related populations and O as the outgroup, this test examines the phylogenetic distribution of derived alleles (designated "B") in P3 at loci that exhibit either an ABBA (ancestral alleles are designated "A") or BABA pattern (see Figure 2a). Under the hypothesis of incomplete lineage sorting, ABBA or BABA sites are expected at equal frequencies, and D-statistic value will be zero. A systematic bias of ABBA or BABA sites, however, would suggest occurrence of introgression between P3 and either P1 (a significantly negative D value) or P2 (a significantly positive D value).

Using the species tree topology of (((East *R. s. sinicus*, Central *R. s. sinicus*), *R. s. septentrionalis*), *R. rex*), this analysis was used to test for introgression between *R. s. septentrionalis* and either East or Central *R. s. sinicus* (see Figure 2b). For this analysis, clean reads from each of the three ingroup taxa were mapped to the gene sequences of the related species *R. rex* (in total 3548 genes, see Results) using BWA-0.5.7 (Li & Durbin, 2009) with default parameters. SAMtools v0.1.19 (Li *et al*., 2009) was used to convert sam files to sorted bam files. SNPs were then called using SAMtools mpileup and bcftools pipeline. We filtered indels, SNPs with read depth < 10, shared SNPs among the three focal taxa, and singleton SNPs in each of the three focal taxa individually. To limit linkage disequilibrium between genes, only one SNP was sampled per gene. This resulted in 1836 SNPs for further analysis.

D-statistic, Z scores and P values were calculated in R version 3.0.1 using the scripts from the evobiR package (http://cran.r-project.org/web/packages/evobiR/index.html). To assess the confidence of the D-statistic in the comparison, we ran 1000 bootstrap pseudoreplicates following previous studies (Eaton & Ree, 2013; Streicher *et al*., 2014).

For our second test of introgression, we investigated alternative bipartitions (excluding the dominant tree) revealed in our earlier BUCKy analysis. To identify partitions, we used CFs of >0.1 following Cui *et al*., (2013). Where CFs show non-overlapping credibility intervals (CIs), the scenario of hybridization can be inferred compared to incomplete lineage sorting (ILS) (see BUCKy tutorial and Ané *et al*., 2007).

Finally, we ran approximately unbiased (AU) tests to determine the number of genes significantly supporting each of three alternative phylogenetic hypotheses that differed in the relationship among the three focal taxa. We defined these topologies as 'resolved' trees: A (((Central *R. s. sinicus*, East *R. s. sinicus*), *R. s. septentrionalis*),

outgroups), B (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), outgroups) and C (((East *R. s. sinicus*, *R. s. septentrionalis*), Central *R. s. sinicus*), outgroups), with the first of these (A) inferred as the species tree. Discordance between the species tree and the other topologies can stem from ILS and/or hybridization. If the discordance is caused by stochastic lineage sorting, the number of genes significantly supporting each of the two non-species tree topologies could be similar, whereas an asymmetry in support for one may suggest hybridization. For this analysis, RAxML was used to generate a file of site-wise log likelihood values for each gene under the three constrained topologies above, and these were then used to generate P values (AU test) in Consel 0.20 (Shimodaira & Hasegawa, 2001).

GENOME DISTRIBUTION OF GENES SUPPORTING THREE DIFFERENT TREE TOPOLOGIES

Following recent species hybridization, introgressed loci might be expected to be more clumped in the genome than they would under a scenario of long-term incomplete lineage sorting. Therefore, to determine the extent to which loci supporting conflicting topologies were clumped in the genome, we evaluated their spatial proximity within the horseshoe bat genome as measured by their membership to genomic scaffolds. Currently, there is no available physical map or high quality genome assembly for a horseshoe bat species. As a reference, we used the fruit bat *Pteropus alecto* (GenBank accession: GCA_000325575.1), which also belongs to suborder Yinpterochiroptera and is expected to show broad synteny with the horseshoe bats. The *P. alecto* genome consists of 65,598 scaffolds with an N50 of 15,954,802 bp. To obtain genomic coordinates of the 3,548 genes, we used the CDS sequences of the *R. s. septentrionalis* as queries in a BLAST against *P. alecto* genomic scaffolds. For each CDS, we identified the scaffold, as well as the gene start and end within its contig.

RESULTS

RNA-SEQ, TRANSCRIPT ASSEMBLY AND ORTHOLOGOUS GENES IDENTIFICATION

We generated RNA-seq data for six horseshoe bat species, yielding an average of 19,861,887 clean paired-end reads per sample (19,265,977-20,684,688 reads; Table 1). Short read data were next assembled *de novo* into 127,096-184,364 transcripts per species. Summary statistics for each individual assembly are provided in Table 2. Transcripts with an average per-site coverage <10x and a sequence length of < 200bp were discarded, leaving approximately 62,316 transcripts per individual with an average length of 1,265 bp.

To identify orthologs, we next conducted a series of tblastx searches and used custom Perl scripts to retrieve the predicted orthologous transcripts from each species data assembly. In total, we identified ~7,590 othologous transcripts per individual with an average gene length of 1,231 bp. Of these, 5,249 transcripts (hereafter called 'genes') were shared by all six horseshoe bat transcriptomes. Genes that did not contain an intact ORF or showed no variation among the six *Rhinolophus* taxa were discarded, leading to a final set of 3,548 genes. For these, we also retrieved the gene sequences from two non-horseshoe bats, *Myotis lucifugus* and *Pteropus vampyrus*, and generated multiple sequence alignments for further analyses. All 13 mitochondrial protein-coding genes were obtained in full from each taxon.

CYTONUCLEAR DISCORDANCE

Maximum Likelihood (ML) phylogenomic reconstruction based on the concatenated nucleotide dataset of 3,548 nuclear genes (3,363,248 bp, 36.1% missing) recovered a fully-resolved topology, with 99% bootstrap probability (BP) supporting the sister-group relationship between Central *R. s. sinicus* and East *R. s. sinicus*, and 100% BP support for this clade's close relationship with *R. s. septentrionalis* (Figure 1b). Notably, *R. rex* was shown to be most closely related to our ingroup taxa. Our second phylogenetic reconstruction approach using BUCKy retrieved an almost identical species phylogeny to the ML tree (see Figure 3 and Figure 1b). Despite the fact that both methods recovered the same tree topology, we found that the concordance factor (CF) for the clade [Central *R. s. sinicus* + East *R. s. sinicus*] was only 0.421 (95% CIs 0.398-0.444), indicating only 42.1% of genes examined in this analysis significantly supported the cluster of Central *R. s. sinicus* and East *R. s. sinicus*. Because there was no difference for concordance factor (CF) among different α values, here we only present results from $\alpha = 1$.

In direct conflict with the species tree inferred from nuclear markers (above), the ML phylogenetic reconstruction based on the concatenated dataset of 13 mitochondrial protein-coding genes (11379 bp, 0.05% missing) recovered a different tree, which strongly supported a sister relationship between East *R. s. sinicus* with the clade uniting Central *R. s. sinicus* and *R. s. septentrionalis* (all BP=100%) (Figure 1c), suggesting *R. s. sinicus* is paraphyletic.

TESTS FOR INTROGRESSION

We used three different methods to assess the incidence of hybridization and introgression among the three focal lineages. First, we performed ABBA-BABA tests of gene flow versus ILS, and found that the D-statistic for the topology given by (((East *R. s. sinicus*, Central *R. s. sinicus*), *R. s. septentrionalis*), *R. rex*) was positive and significantly greater than zero (D = 0.1327, P = 1.6×10^{-5}). This result indicates a systematic bias of BABA sites shared by *R. s. septentrionalis* (P3) and Central *R. s. sinicus* (P2) (see Figure 2b), suggesting introgression between Central *R. s. sinicus* and *R. s. septentrionalis*, but no such gene flow between East *R. s. sinicus* and *R. s. septentrionalis.* In order to rule out the possible effect of selection loci on the ABBA-BABA tests, we tested for positive selection using site models in the PAML package version 4.9 (Yang, 2007) and identified 23 genes that showed evidence of positive selection (models M7 vs M8). Bayes empirical Bayes (BEB) analyses of these genes found 30 sites with a posterior probability of >0.95, none of which were among the 1836 sites used in the ABBA-BABA test (data not shown).

Second, the BUCKy analysis revealed two bipartitions with CFs of >0.1 that corresponded to topologies discordant with the species tree. The first of these alternative topologies contained the clade [Central *R. s. sinicus* + *R. s. septentrionalis*] with a CF of 0.363 (95% credibility intervals 0.341-0.385), and the second contained the clade [East *R. s. sinicus* + *R. s. septentrionalis*] with a CF of 0.215 (95% CIs 0.196-0.235). These non-overlapping credibility intervals may be indicative of hybridization.

Finally, an asymmetry in support for one of the two non-species tree topologies (i.e. (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), outgroups)) suggested hybridization between Central *R. s. sinicus* and *R. s. septentrionalis* (see Figure 4). Specifically, among 1065 genes (~30% of 3548 genes examined) that gave well-supported tree topologies (AU test), there were similar numbers of genes supporting the species tree topology (392 genes) and the topology of (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), outgroups) (408 genes), suggestive of a mosaic genome for Central *R. s. sinicus* derived from both *R. s. septentrionalis* and East *R. s. sinicus*.

GENOME DISTRIBUTION OF GENES SUPPORTING THREE DIFFERENT TREE TOPOLOGIES

To examine whether the genes supporting each of the three conflicting tree topologies were clumped in the genome, we mapped all 3,548 genes against the genome of the fruit bat *P. alecto*. We successfully obtained the genomic coordinates for 3,262 of our genes, which were found spread across 247 genomic scaffolds in *Pteropus* genome. In total, 1,065 genes showed a strong signal of support for one of the three topologies, and these were mapped to 187 scaffolds. Genes showing a clear phylogenetic signal for the species tree (((East *R. s. sinicus*, Central *R. s. sinicus*), *R. s. septentrionalis*), outgroups) were distributed across 138 scaffolds (N=392), while genes supporting the two alternative hypotheses, i.e. (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), outgroups) and (((East *R. s. sinicus*, *R. s. septentrionalis*), Central *R. s. sinicus*), outgroups) were found in 131 and 114 scaffolds, respectively. To test further for genomic proximity among loci supporting particular phylogenetic hypotheses, we examined levels of synteny for those CDSs that mapped to the same scaffold. Of the 187 scaffolds containing genes, 11 contained just one and were thus uninformative,

118 contained genes with signal for at least two alternative phylogenetic scenarios, and 78 scaffolds contained genes with mixed support for all three hypotheses tested. Scaffolds containing genes supporting a single type of tree were equally found for all three topologies tested: 29 scaffolds for the species tree, 23 for (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), outgroups) and 17 scaffolds with only genes supporting the (((East *R. s. sinicus*, *R. s. septentrionalis*), Central *R. s. sinicus*), outgroups) topology. Consequently, we found no clear clustering of genes supporting a given topology. This scatter distribution pattern of genes supporting different tree topologies indicated that as a null hypothesis incomplete lineage sorting could not be ruled out.

DISCUSSION

Convincing cases of nuclear introgression in vertebrate groups are rare, in part due to the problems of ruling out incomplete lineage sorting, and also because of the limited power of the genetic markers that have been used in most studies to date (e.g. Zieliński *et al*., 2013; Pons *et al*., 2014). High throughout sequencing offers new opportunities and approaches for detecting signatures of introgression by sampling loci more widely across genomes (Twyford & Ennos, 2012). Here we used RNAseq data to obtain sequence data from 3,548 loci from three closely related forms of the Chinese horseshoe bat *Rhinolophus sinicus*, and used these to diagnose secondary ncDNA introgression against an expected background of incomplete lineage sorting in the context of secondary contact.

Our phylogenomic reconstructions using ML and Bayesian concordance analysis strongly supported a sister relationship between *R. s. septentrionalis* and the clade containing East and Central *R. s. sinicus*, so corroborating earlier morphological analyses (Mao *et al*., 2013b). This inferred species tree, however, contradicted the one obtained from the complete mitogenome, which instead supported a sister relationship between Central *R. s. sinicus* and *R. s. septentrionalis*, with East *R. s. sinicus* placed outside of this clade. This phylogenetic discordance was previously reported based on analyses of two mitochondrial genes and four nuclear loci (Mao *et al*., 2013b), which, together with estimates of gene flow obtained from isolation-with-migration (IM) models, was interpreted as evidence of directional mtDNA introgression from *R. s. septentrionalis* to Central *R. s. sinicus*, implying hybridization between females of the former taxon with males of the latter (Mao *et al*., 2013b).

Yet contrary to our earlier work that found no evidence of ncDNA gene flow between the focal taxa (Mao *et al*., 2013b), new analyses of >3500 loci with three different methods (BUCKy, AU test, and D-statistic test) strongly suggests that ncDNA introgression has taken place. This discrepancy almost certainly reflects the much greater genetic sampling in our new study, and thus highlights the benefits of genomic approaches for elucidating patterns of divergence and gene flow (e.g. Stölting *et al*., 2013; Palmer & Kronforst, 2015). In particular, we found that coding genes of the sequenced individual of Central *R. s. sinicus* grouped at broadly equally frequencies with loci from *R. s. septentrionalis* and East *R. s. sinicus*. Formal ABBA-BABA tests applied to the large set of gene trees suggested that levels of allele sharing between *R. s. septentrionalis* and Central *R. s. sinicus* were significantly greater than would be expected under a scenario of ILS acting alone (see Durand *et al*., 2011). Such a result is unlikely to stem from the presence of ghost lineages, which can sometimes lead to erroneous inferences of gene flow in ABBA-BABA tests (see Durand *et al*., 2011; Eaton & Ree, 2013; Rogers & Bohlender, 2015; Eaton *et al*., 2015). Indeed, geographical sampling from the study region has been extensive, and to date has not revealed the presence of additional closely related taxa on the mainland of China (Mao et al. 2013b). In addition, although samples of *R. sinicus* from outside of China (e.g. the Himalayas) have not been studied, preliminary analysis of mitochondrial *cytochrome b* sequences (Stoffberg *et al*., 2010) showed marked divergence between a sample from Nepal and our three focal taxa, indicating that this geographically isolated lineage from the Himalayas would have no effect on the ABBA-BABA tests here.

Despite the evidence for ncDNA introgression between Central *R. s. sinicus* and *R. s. septentrionalis*, it is important to recognise that AU and D-statistic tests do not quantify the relative contribution of ILS and introgression, and thus cannot rule out ILS. Instead, in each of these two tests, introgression is accepted over a null hypothesis of ILS occurring alone, whereas in most real cases of recent divergence histories, these processes will almost certainly not be mutually exclusive (e.g. Liu *et al*., 2014; Marcussen *et al*., 2014).

An alternative explanation for the pattern of genomic mosaicism observed in Central *R. s. sinicus* is that it is a hybrid species, formed by interbreeding between East *R. s. sinicus* males and *R. s. septentrionalis* females. To date, there are several reported examples of homoploid hybrid speciation in animals (e.g. insects, Gompert *et al*., 2006; fishes, Cui *et al*., 2013; birds, Brelsford *et al*., 2011; also see reviews in Mavárez & Linares, 2008 and Schumer *et al*., 2014) including one case from bats (Larsen *et al*., 2010). Yet we consider that hybrid lineage is an implausible scenarios in our study system. Earlier analyses provided compelling evidence that the two geographical clades of *R. s. sinicus* diverged after their common ancestor split with *R. s. septentrionalis*; consequently the order of these events is inconsistent with a hybrid species origin, and instead genomic mosacism is more likely a consequence of the combined effects of ILS and secondary introgression. More generally, caution is required when accepting scenarios of hybrid speciation, especially when genetic sampling is limited, because different loci in the genome may vary in their degree of introgression across species boundaries (Hamilton *et al*., 2013; Larson *et al*., 2014; Harrison & Larson, 2016). A recent example comes from a study focussing on swordtails of the genus *Xiphophorus*, in which a hypothesis of hybrid speciation for *X. clemenciae* (Meyer *et al*., 2006) has since been ruled out following genome-wide sequencing (Schumer *et al*., 2013).

To test further for ILS and secondary introgression in producing genomic mosaicism in Central *R. s. sinicus*, we also attempted to examine the distribution of loci supporting each of the three tree topologies (see also Martinsen *et al*., 2001; Schumer *et al*., 2013). We found that genes supporting each of the three tree topologies were not clustered and thus ILS cannot be ruled out for the observed gene conflict within this species complex (see also Pollard *et al*., 2006), although ancient hybridization and subsequent recombination will also lead to similarly scattered distributions (see Schumer *et al*., 2013). It is also unclear whether inferences of genomic distributions in horseshoe bats based on the reference genome of *Pteropus alecto* are reliable given that these taxa, although belonging to the same suborder, might have different genomic architectures. As more and better quality bat genomes become available, we will be able to assess whether such assumptions of gene synteny are safe.

Although genome sequencing costs have plummeted in the past few years, genomic library construction remains relatively expensive, reducing the affordability of genomics studies of many individuals and populations. Thus a trade-off between sequencing small numbers of loci in many individuals, versus sequencing large numbers of loci in few individuals, still persists. Landguth *et al*.,. (2012) suggested that increasing the power of detecting the effects of landscape patterns on gene flow might be better achieved by studying more loci than by increasing numbers of individuals, while Robinson *et al*., (2014) made similar recommendations in the context of improved model selection and parameter estimation in approximate

Bayesian computation (ABC). Other recent studies have also studied gene flow and admixture based on transciptomes or low coverage genomes of single individuals sampled from populations or species (e.g. Schumer *et al*., 2014; Frantz *et al*., 2013 and 2014; Hearn *et al*., 2014; Gante *et al*., 2016) or transcriptome sequences (e.g. Cui *et al*., 2013). Therefore, while population genomics studies will inevitably become increasingly common in the coming years, studies of gene flow and divergence such as ours, based on deep genomic coverage of small numbers of individuals, can still make useful contributions.

CONCLUSIONS

We used genome-wide data to elucidate the phylogenetic relationships among a species complex of horseshoe bats, Central *R. s. sinicus*, East *R. s. sinicus* and *R. s. septentrionalis*. Phylogenetic analyses of 3,548 genes provided firm support for the evolutionary relationships of the three recently diverged taxa, but also extensive genomic mosaicism in Central *R. s. sinicus* genome that has arisen from secondary introgression and/or ILS between this taxon and *R. s. septentrionalis* in their evolutionary history. Our analyses did not find any evidence for a hybrid lineage in this model system, although further analyses of the extent and timing of hybridization and introgression may allow us to disentangle the relative contributions of ILS and introgression in shaping the genomes of these species (Payseur & Rieseberg, 2016). Finally, our study further highlights that horseshoe bats can be an excellent model system for studying speciation, and further population genomic surveys for this species complex could allow us to identify genes involved in reproductive isolation between recently diverged taxa, as well as to test for adaptation associated with secondary introgression in natural populations.

ACKNOWLEDGEMENTS

We thank Joe Parker and Kalina Davies for help in analysing the data, and we also thank Zhang Junpeng and Zhu Guangjian for assistance with fieldwork. We also thank four anonymous reviewers for their helpful comments that improved an earlier version of the manuscript. X Mao was supported by a Marie Curie International Incoming Fellowship, a National Natural Science Foundation of China grant (No. 31570378), and the Shanghai Pujiang Talent Program Foundation, and S. Rossiter was supported by a European Research Council Starting Grant (No. 310482).

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JB, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246.

Alvarez M, Schrey AW, Richards CL. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology* **24**: 710-725.

Anés C, Larget B, Baum DA, Smith SD, Rokas A. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* **24**: 412-426.

Berger SA, Krompass D, Stamatakis A. 2011. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology* **60**: 291–302.

Berthier P, Excoffier L, Ruedi M. 2006. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B-Biological Sciences* **273**: 3101-3109.

Bogdanowicz W, Piksa K, Tereba A. 2012. Hybridization hotspots at bat swarming sites. *PLoS ONE* **7**: e53334.

Bogdanowicz W, Van Den Bussche RA, Gajewska M, Postawa T, Harutyunyan M. 2009. Ancient and contemporary DNA sheds light on the history of mouse-eared bats in Europe and the Caucasus. *Acta Chiropterologica* **11**: 289-305.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. 2014. Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PloS Genetics* **10**: e1004410.

Brelsford A, Milá B, Irwin DE. 2011. Hybrid origin of Audubon's warbler. *Molecular Ecology* **20**: 2380-2389.

Chattopadhyay B, Garg KM, Kumar AKV, Doss DPS, Rheindt FE, Kandula S, Ramakrishnan U. 2016. Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. *BMC Evolutionary Biology* **16**: 41.

Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, WebsterM, Antão T, MacInnis B, Kwiatkowski D, Donnelly MJ. 2014. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications* **5**:4248.

Clark AG. 1997. Neutral behavior of shared polymorphism. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 7730–7734.

Csorba G, Ujhelyi P, Thomas N. 2003. Horseshoe Bats of the World (Chiroptera: Rhinolophidae). Alana Books, Shropshire, U. K 118 Rouge.

Cui RF, Schumer M, Kruesi K, Walter R, Andolfatto P Rosenthal GG. 2013. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**: 2166-2179.

Dai M, Thompson RC, Maher C, Contreras-Galindo R, Kaplan MH, Markovitz DM, Omenn G, Meng F. 2010. NGSQC: cross-platform quality analysis pipeline for deep sequencing data. *BMC Genomics* **11**: S7.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* **28**: 2239–2252.

Eaton DA, Hipp AL, González-Rodríguez A, Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**: 2587-2601.

Eaton DA, Ree RH. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* **62**: 689-706.

Fernández R, Laumer CE, Vahtera V, Libro S, Kaluziak S, Sharma PP, Pérez-Porro AR, Edgecombe GD, Giribet G. 2014. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Molecular Biology and Evolution* **31**: 1500-1533.

Frantz LAF, Madsen O, Megens HJ, Groenen MA, Lohse K. 2014. Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian *Sus* species during the Plio-Pleistocene climatic fluctuations. *Molecular Ecology* **23**:5566-5574.

Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Bosse M, Paudel Y, Semiadi G, Meijaard E, Li N, Crooijmans RP, Archibald AL, Slatkin M, Schook LB,Larson G, Groenen MA. 2013. Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biology* **14**: R107.

Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular Ecology* (in press).

Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC. 2006. Homoploid hybrid speciation in an extreme habitat. *Science* **314**: 1923–1925.

Grabherr MG, Aubert G, Carrère S, Cruaud C, Brochot AL, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J, da Silva C, Moreau S,Gamas P, Wincker P, Gouzy J, Burstin J. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644-652.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD,Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A,Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M,Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722.

Hamilton JA, Lexer C, Aitken SN. 2013. Differential trogression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone. *New Phytologist* **197**: 927-938.

Harrison RG, Larson EL. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology* **25**: 2454-2466.

Hawkins RD, Hon GC, Ren B. 2010. Next-generation genomics: an integrative approach. *Nature Reviews Genetics* **11**: 476-486.

Hearn J, Stone GN, Bunnefeld L, Nicholls JA, Barton NH, Lohse K. 2014. Likelihood-based inference of population history from low-coverage *de novo* genome assemblies. *Molecular Ecology* **23**: 198-211.

Hodgins KA, Lai Z, Oliveira LO, Still DW, Scascitelli M, Barker MS, Kane NC, Dempewolf H, Kozik A, Kesseli RV, Burke JM, Michelmore RW, Rieseberg LH. 2014. Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molecular Ecology Resources* **14**: 166-177.

Hoffmann FG, Owen JG, Baker RJ. 2003. mtDNA perspective of chromosomal diversification and hybridization in Peters′ tent-making bat (*Uroderma bilobatum*: Phyllostomidae). *Molecular Ecology* **12**: 2981-2993.

Hudson ME. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**: 3–17.

Katoh K, Kuma K, Toh H, Miyata T. 2007. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511-518.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**: 17-24.

Landguth EL, Fedy BC, Oyler-Mccance SJ, Garey A, Emel SL, Mumma M, Wagner HH, Fortin MJ, Cushman SA. 2012. Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology* Resources **12**: 276-284.

Larget B, Kotha SK, Dewey CN, An é C. 2010. BUCKy: Gene tree/species tree reconciliation with the Bayesian concordance analysis. *Bioinformatics* **26**: 2910-2911.

Larsen PA, Marcha ń-Rivadeneira MR, Baker RJ. 2010. Natural hybridization generates mammalian lineage with species characteristics. *Proceedings of the National Academy of Sciences, USA* **107**: 11447–11452.

Larson EL, Andres JA, Bogdanowicz SM, Harrison RG. 2014. Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution* **67**: 3653-3661.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liu K, Steinberg E, Yozzo A, Song Y, Kohn M, Nakhleh L. 2014. Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences of the USA* **112**: 196–201.

Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **363**:2971–2986.

Mao XG, Zhang JP, Zhang SY, Rossiter SJ. 2010. Historical male-mediated introgression in horseshoe bats revealed by multi-locus DNA sequence data. *Molecular Ecology* **19**: 1352-1366.

Mao XG, He GM, Hua PY, Jones G, Zhang SY, Rossiter SJ. 2013a. Historical introgression and the persistence of alleles in the intermediate horseshoe bat (*Rhinolophus affinis*). *Molecular Ecology* **22**:1035-1050.

Mao XG, He GM, Zhang JP, Rossiter SJ, Zhang SY. 2013b. Lineage divergence and historical gene flow in Chinese horseshoe bat (*Rhinolophus sinicus*). *PLoS One* **8**: e5678.

Mao XG, Thong VD, Bates PJJ, Jones G, Zhang SY, Rossiter SJ. 2013c. Multiple cases of asymmetric introgression among horseshoe bats detected by phylogenetic conflicts across loci. *Biological Journal of the Linnean Society* **110**: 346-361.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, Olsen O. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092.

Marra NJ, Romero A, DeWoody JA. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. *Molecular Ecology* **23**: 2699-2711.

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.

Mavárez J, Linares M. 2008. Homoploid hybrid speciation in animals. *Molecular Ecology* **17**: 4181-4185.

Meyer A, Salzburger W, Schartl M. 2006. Hybrid origin of a swordtail species(Teleostei: *Xiphophorus clemenciae*) driven by sexual selection. *Molecular Ecology* **15**: 721–730.

Neubaum M, Douglas MR, Douglas ME, O'shea TJ. 2007. Molecular ecology of the big brown bat (*Eptesicus fuscus*): genetic and natural history variation in a hybrid zone. *Journal of Mammalogy* **88**: 1230-1238.

Nolte A, Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends in Genetics* **26**: 54-58.

Norrisa LC, Maina BJ, Leea Y, Colliera TC, Fofanad A, Cornela AJ, Lanzaroa GC. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences, USA* **112**: 815–820.

Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WE, Nickel M, Schierwater B, Vacelet J, Wiens M, Wörheide G. 2013. Deep metazoan phylogeny: when different genes tell different stories. *Molecular Phylogenetics and Evolution* **67**: 223–233.

Palmer DH, Kronforst MR. 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.

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology* **25**: 2337-2360.

Pereira RJ, Barreto FS, Pierce T, Carneiro M, Burton RS. 2016. Transcriptome-wide patterns of divergence during allopatric evolution. *Molecular Ecology* **25**: 1478–1493.

Phifer-Rixey M, Bomhoff M, Nachman M. 2014. Genome-wide patterns of differentiation among house mouse subspecies. *Genetics* **198**: 283-297.

Pollard DA, Lyer VN, Moses AM, Eisen MB. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence from incomplete lineage sorting. *PLoS Genetics* **2**: e173.

Pon J-M, Sonsthagen S, Dove C, Crochet P-A. 2014. Extensive mitochondrial introgression in North American Great Black-backed Gulls (*Larus marinus*) from the American Herring Gull (*Larus smithsonianus*) with little nuclear DNA impact. *Heredity* **112**: 226-239.

Robinson J, Bunnefeld L, Hearn J, Stone GN, Hickerson MJ. 2014. ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology* **23**: 4458-4471.

Rogers, A. R., and R. J. Bohlender. 2015. Bias in estimators of archaic admixture. *Theoretical Population Biology* **100**: 63–78.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Salichos, L. & Rokas, A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**: 327–331.

Schumer M, Cui RF, Boussau B, Walter R, Rosenthal G. 2013. An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. *Evolution* **67**: 1155-1168.

Schumer M, Rosenthal G, Andolfatto P. 2014. How common is homoploid hybrid speciation? *Evolution* **68**: 1553-1560.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**: 198–207.

Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, giribet G. 2014. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution* **31**: 2963-2984.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* **26**: 1135-1145.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**:1246-1247.

Song Y, Endepols S, Klemann N, Richter D, Matuschka F, Shih C, Nachman MW, Kohn MH. 2011. Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. *Current Biology* **21**: 1296-1301.

Stoffberg S, Jacobs DS, Mackie IJ, Matthee CA. 2010. Molecular phylogenetics and historical biogeography of *Rhinolophus* bats. *Molecular Phylogenetics and Evolution* **54**: 1-9.

Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C. 2012. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**: 842-855.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK. 2014. Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology* **23**: 3273-3291.

Stuglik MT, Babik W. 2016. Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecology and Evolution* **6**: 4513-4525.

Tepolt CK, Palumbi SR. 2015. Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology* **24**: 4145-4158.

Twyford AD, Ennos RA. 2012. Next-generation hybridization and introgression. *Heredity* **108**: 179-189.

Vallo P, Benda P, Červený J, Koubek P. 2013. Conflicting mitochondrial and nuclear paraphyly in small-sized west African house bats (Vespertilionidae). *Zoological Scrience* **42**:1-12.

Webb NJ, Tidemann CR. 1995. Hybridization between black (*Pteropus alecto*) and grey-headed (*P. poliocephalus*) flying-foxes (Megachiroptera: Pteropodidae). *Australian Mammalogy* **18**:19-26.

Weisrock DW, Smith SD, Chan LM, Biebouw K, Kappeler PM, Yoder AD. 2012. Concatenation and concordance in the reconstruction of mouse lemur phylogeny:

an empirical demonstration of the effect of allele sampling in phylogenetics. *Molecular Biology and Evolution* **29**: 1615-1630.

Westram AM, Galindo J, Rosenblad MA, Grahame JW, Panova M, Butlin RK. 2014. Do the same genes underlie parallel phenotypic divergence in different Littorina saxatilis populations? *Molecular Ecology* **23**: 4603-4616.

Xu WX, Zhou ZM, Zhang JQ, Wu Y, Li YC, Hu JC. 2005. Study on geographical variation of skull morphology of *Rhinolophus sinicus*. *Sichuan Journal of Zoology* **24**: 469-472.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.

Zhang W, Kunte K, Kronforst MR. 2013. Genome-wide characterization of adaptation and speciation in tiger swallowtail butterflies using de novo transcriptome assemblies. *Genome Biology and Evolution* **5**: 1233-1245.

Zieliński P, Nadachowska-Brzyska K, Wielstra B, Szkotak R, Covaciu-Marcov SD, Cogălniceanu D, Babik W. 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Molecular Ecology* **22**: 1884-1903.

Figure legends

Figure 1. Sampling map and ML trees reconstructed based on nuclear and mitochondrial genomes, respectively. (a) Map showing distribution ranges of the three *R. sinicus* taxa, Central *R. s. sinicus*, East *R. s. sinicus* and *R. s. septentrionalis* (modified from Mao *et al* 2013). Red arrows indicate the sampling sites of bats in this study; (b) ML-tree based on a concatenated dataset of 3548 nuclear protein-coding genes; (c) ML-tree based on a concatenated dataset of 13 mitochondrial protein-coding genes

Figure 2. Analysis of introgression based on the ABBA-BABA test. (a) An example of population comparison used in the test; (b) three comparisons of our focal taxa used in the test. D1 (((Central *R. s. sinicus*, East *R. s. sinicus*), *R. s. septentrionalis*), *R. rex*), D2 (((Central *R. s. sinicus*, *R. s. septentrionalis*), East *R. s. sinicus*), *R. rex*) and D3 (((East *R. s. sinicus*, *R. s. septentrionalis*), Central *R. s. sinicus*), *R. rex*).

Figure 3. A primary concordance tree generated by BUCKy from 2919 gene trees. Nodal values are Bayesian concordance factors and their credibility intervals (given above and below the branch, respectively).

Figure 4. Number of genes supporting each of the three tree topologies among the three focal taxa and the outgroups.

Table 1 Statistics of clean RNA-seq data.

Samples	Total reads	Total nucleotides (nt)	Q20 percentage(%)	GC percentage(%)
East *R. s. sinicus*	19,408,192	3,493,474,560	97.83	49.13
Central *R. s. sinicus*	20,082,158	3,614,788,440	97.8	49.67
R. s. septentrionalis	19,523,461	3,514,222,980	97.78	50.8
R. rex	20,684,688	3,723,243,840	98.28	49.9
R. pearsoni chinensis	20,206,846	3,637,232,280	98.37	50.62
R. ferrumequinum	19,265,977	3,467,875,860	98.25	50.38

865 Table 2 Summary of the assembled results by Trinity.

Samples	N^a	N^b	Average length (bp)	N^c	Average length (bp)
East *R. s. sinicus*	184,364	73,678	1290	7,868	1,274
Central *R. s. sinicus*	127,096	51,243	1167	7,266	1,178
R. s. septentrionalis	134,026	55,019	1182	6,804	1,164
R. rex	179,485	71,615	1285	7,738	1,240
R. pearsoni chinensis	152,341	64,968	1470	8,319	1,355
R. ferrumequinum	135,796	57,372	1201	7,542	1,174

N^a: Number of assembled contigs; N^b: Number of contigs (>=10x and >=200bp); N^c: Number of genes

870 Figure 1

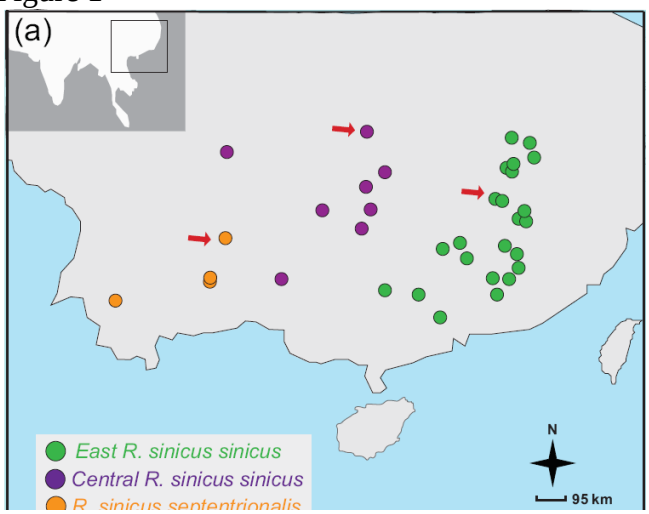
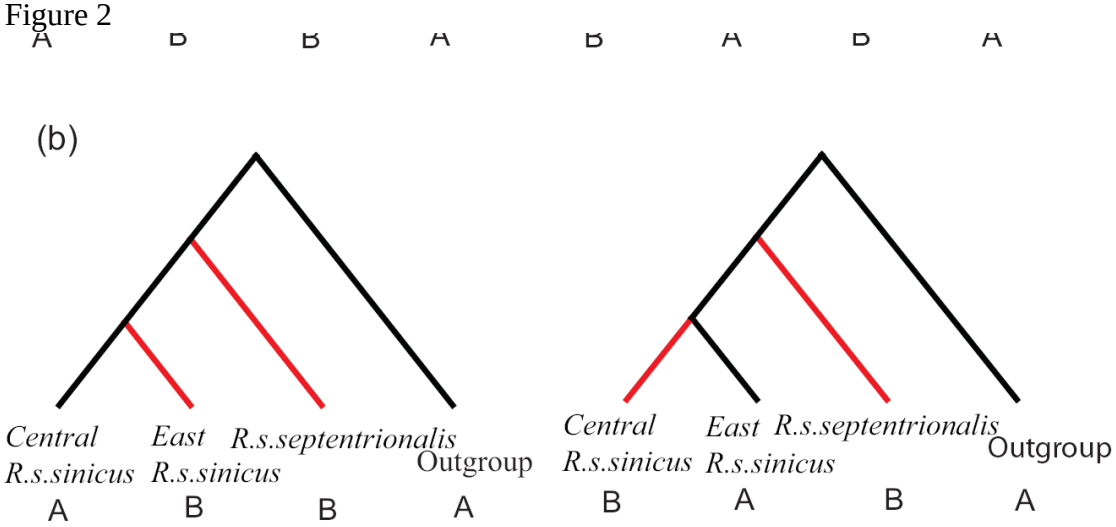


Figure 2


(b)
Central
R.s.sinicus
East
R.s.sinicus
R.s.septentrionalis
Outgroup
A
B
B
A
Central
R.s.sinicus
East
R.s.sinicus
R.s.septentrionalis
Outgroup
B
A
B
A

875

Figure 3

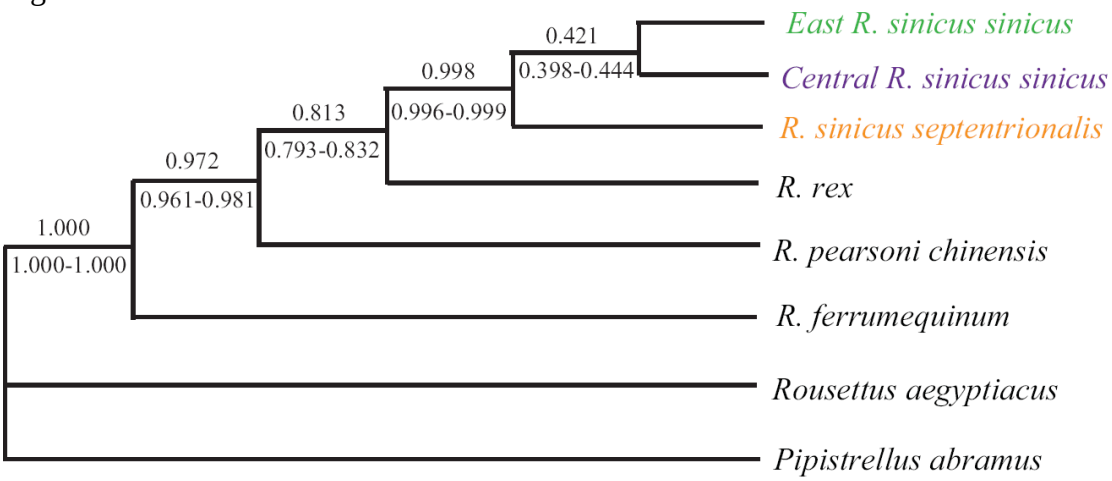
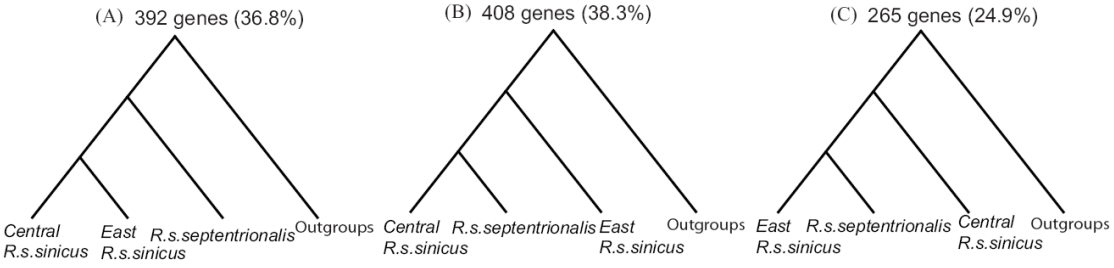

0.421
0.398-0.444
0.998
0.996-0.999
0.813
0.793-0.832
0.972
0.961-0.981
1.000
1.000-1.000
East R. sinicus sinicus
Central R. sinicus sinicus
R. sinicus septentrionalis
R. rex
R. pearsoni chinensis
R. ferrumequinum
Rousettus aegyptiacus
Pipistrellus abramus

Figure 4


(A) 392 genes (36.8%)
Central R.s.sinicus
East R.s.sinicus
R.s.septentrionalis
Outgroups
(B) 408 genes (38.3%)
Central R.s.sinicus
R.s.septentrionalis
East R.s.sinicus
Outgroups
(C) 265 genes (24.9%)
East R.s.sinicus
R.s.septentrionalis
Central R.s.sinicus
Outgroups

880