MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 668–688 doi: 10.1111/mec.13920

INVITED REVIEWS AND SYNTHESES

# Patterns of MHC-dependent mate selection in humans and nonhuman primates: a meta-analysis

J. WINTERNITZ,*† ‡ J. L. ABBATE,§¶ E. HUCHARD,** J. HAVLÍČEK† † and L. Z. GARAMSZEGI‡ ‡

*Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, August-Thienemann-Strasse 2, 24306 Ploen, Germany, †Institute of Vertebrate Biology, Czech Academy of Sciences, v.v.i., Květná 8, 603 65 Brno, Czech Republic, ‡Institute of Botany, Czech Academy of Sciences, v.v.i., Lidická 25/27, 657 20 Brno, Czech Republic, §Institute of Ecology and Evolution, University of Bern, Balterstrasse 6, 3006 Bern, Switzerland, ¶INRA - UMR 1062 CBGP (INRA, IRD, CIRAD, Montpellier SupAgro), 755 Avenue du campus Agropolis, 34988 Montferrier-sur-Lez, France, **CEFE UMR5175, CNRS - Université de Montpellier - EPHE, 1919 Route de Mende, 34295 Montpellier Cedex 5, France, ††Department of Zoology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague 2, Czech Republic, ‡‡Department of Evolutionary Ecology, Estación Biológica de Doñana-CSIC, c/Americo Vespucio, s/n, 41092, Seville, Spain

## Abstract

**Genes of the major histocompatibility complex (MHC) in vertebrates are integral for effective adaptive immune response and are associated with sexual selection. Evidence from a range of vertebrates supports MHC-based preference for diverse and dissimilar mating partners, but evidence from human mate choice studies has been disparate and controversial. Methodologies and sampling peculiarities specific to human studies make it difficult to know whether wide discrepancies in results among human populations are real or artefact. To better understand what processes may affect MHC-mediated mate choice across humans and nonhuman primates, we performed phylogenetically controlled meta-analyses using 58 effect sizes from 30 studies across seven primate species. Primates showed a general trend favouring more MHC-diverse mates, which was statistically significant for humans. In contrast, there was no tendency for MHC-dissimilar mate choice, and for humans, we observed effect sizes indicating selection of both MHC-dissimilar and MHC-similar mates. Focusing on MHC-similar effect sizes only, we found evidence that preference for MHC similarity was an artefact of population ethnic heterogeneity in observational studies but not among experimental studies with more control over sociocultural biases. This suggests that human assortative mating biases may be responsible for some patterns of MHC-based mate choice. Additionally, the overall effect sizes of primate MHC-based mating preferences are relatively weak (Fisher's Z correlation coefficient for dissimilarity Zr = 0.044, diversity Zr = 0.153), calling for careful sampling design in future studies. Overall, our results indicate that preference for more MHC-diverse mates is significant for humans and likely conserved across primates.**

*Keywords*: good genes, HLA, inbreeding avoidance, major histocompatibility complex, mating preference, sexual selection

*Received 18 May 2016; revision received 14 October 2016; accepted 19 October 2016*

## Introduction

The major histocompatibility complex (MHC), an extraordinarily polymorphic and ancient chromosomal region shared by virtually all vertebrates, is the most likely candidate for 'good genes' – genes with fitness benefits – due to its involvement in both immune defence and mate choice (Potts & Wakeland 1990; Hedrick 1994; Bernatchez & Landry 2003; Milinski 2006). The MHC encodes molecules that bind specific self- and pathogen-derived peptides and present these

Correspondence: Jamie Winternitz, Fax: +49 4522 763 310; E-mail: jcwinternitz@gmail.com

© 2016 John Wiley & Sons Ltd

to T lymphocytes, thus initiating appropriate immune activation (Hughes & Yeager 1998). Class I molecules mainly bind intracellular pathogen peptides (e.g. viruses and bacteria) and are expressed by all nucleated cells, whereas class II molecules mainly bind extracellular parasite peptides (e.g. helminths, ectoparasites) and are expressed by professional immune antigen-presenting cells, such as mononuclear phagocytes or T cells (Knapp 2005). Currently over 9000 human MHC (HLA) class I alleles and over 3000 class II alleles have been recorded across human populations (Robinson *et al.* 2014). This hyperpolymorphism is attributed to parasite-mediated balancing selection promoting host allelic diversity to defend against a dynamic spectrum of parasites (Bernatchez & Landry 2003; Wegner *et al.* 2003a; Prugnolle *et al.* 2005; Šimková *et al.* 2006; Goüy de Bellocq *et al.* 2008; Solberg *et al.* 2008; Garamszegi & Nunn 2011; Eizaguirre *et al.* 2012; Garamszegi 2014).

Infectious agents are thought to be the strongest selective force shaping human evolutionary history (McMichael 2001; Fumagalli *et al.* 2011; Karlsson *et al.* 2014) and continue to have strong effects on fitness. For example, humans are currently known to be infected by over 1400 parasitic species (Taylor *et al.* 2001) and in 2010 parasites were responsible for nearly 64% of global deaths in children younger than 5 years (Liu *et al.* 2012). Parasite-mediated balancing selection in humans is identifiable at both broad and fine scales. At the population level, parasite-mediated balancing selection is indicated by spatial patterns of MHC polymorphism increasing with virus species richness (Prugnolle *et al.* 2005) and greater frequencies of protective MHC alleles in areas with greater parasite risk (Hill *et al.* 1991). Selection at the individual level is indicated by resistance of rare MHC genotypes to specific strains of pathogens (Thursz *et al.* 1995; Trachtenberg *et al.* 2003) with reciprocal selection for pathogen escape mutations to avoid MHC immune control (Goulder *et al.* 2001).

While it is clear that population MHC allelic diversity is associated with pathogen diversity and prevalence, there is also theoretical and empirical support for the role of sexual selection in maintaining heterozygosity and allelic variation across the MHC (Potts *et al.* 1991, 1994; Hedrick 1992; Jordan & Bruford 1998; Penn & Potts 1999; Winternitz *et al.* 2013; Ejsmond *et al.* 2014). Proximate mechanisms enabling MHC-mediated mate choice include odour and visual cues of MHC composition. Animals can discriminate between the different volatile peptides bound by products of specific alleles that contribute to body odours (Boyse *et al.* 1983; Potts *et al.* 1994; Carroll *et al.* 2002; Penn 2002; Leinders-Zufall *et al.* 2004; Milinski *et al.* 2005, 2013). Additionally, visual cues such as the expression of sexually dimorphic conspicuous traits, or even condition-related behaviour, can be indicators of immune genotype (Hamilton & Zuk 1982; Folstad & Karter 1992).

Ultimate mechanisms for MHC-mediated mate choice can take three nonmutually exclusive forms (Piertney & Oliver 2006). (i) Preferences for MHC diversity (measured in terms of heterozygosity, or the number of MHC alleles) in mates could provide direct fitness benefits (e.g. healthier mates are better providers and less infectious to their mates and offspring) and/or indirect benefits if rare alleles, more likely to be carried by heterozygotes or by individuals harbouring many alleles, can be passed on to offspring [e.g. 'good-genes-as-heterozygosity' (Brown 1997, 1999)], although this remains to be demonstrated empirically. Thus, preferences for MHC diversity in mates could also increase offspring levels of heterozygosity in structured, finite populations (Fromhage *et al.* 2009). (ii) Preferences for specific MHC genotypes in potential mates could provide the direct benefits mentioned previously and also indirect benefits if genes that are protective against contemporary parasites can be passed to offspring. Protective alleles are thought to be more rare in the population (Slade & McCallum 1992) and so more likely to be carried by heterozygotes. Both preferences for heterozygosity and preferences for specific resistance genotypes can broadly fit under the category of preferences for MHC diversity for our current study. (iii) Preferences for MHC dissimilarity/complementarity in mates would provide indirect fitness benefits by increasing immunodiversity of offspring (Tregenza & Wedell 2000). Alternatively, preferences for MHC-dissimilar mates may be independent of immunogenetic benefits for offspring if MHC alleles serve as markers of relatedness in mates. This would provide indirect fitness benefits by avoiding consequences of close inbreeding (Yamazaki *et al.* 1988). If preference for dissimilar mates is for immunogenetic indirect benefits, then the extent of dissimilarity of the potential mate may matter. The optimality hypothesis proposes that offspring may benefit most from optimal rather than maximal immunodiversity (Wegner *et al.* 2003b; Milinski 2006), because having too many different alleles could actually deplete autoreactive T cells that are required for immune response (Nowak *et al.* 1992; Woelfing *et al.* 2009). In this case, the degree and composition of optimal MHC diversity for offspring are expected to primarily depend on the diversity of parasites in the environment (Wegner *et al.* 2003b; Milinski 2006; Eizaguirre *et al.* 2009).

Instead of mediating mate choice preferences, the MHC may be incidentally associated with signals of overall condition and individual vigour that are dependent on genomewide heterozygosity (Brown 1997). Similarly, MHC allele frequencies that vary by population may simply correlate with phenotypic cues of genetic

© 2016 John Wiley & Sons Ltd

relatedness. The absence of a correlation between MHC and neutral variation would support MHC-mediated mate choice, although the presence of a correlation would not necessarily preclude the MHC's role.

A recent meta-analysis found broad evidence of MHC-associated mate choice in nonhuman species, with stronger support for diversity preferences than for dissimilarity preferences (Kamiya *et al.* 2014), but results from human studies have been more equivocal and contentious (reviewed in Havlicek & Roberts 2009; Winternitz & Abbate 2015). This is in large part due to the greater variation inherent in human research because many potentially confounding aspects of the study design are harder to control. The most problematic issue is likely hidden, yet substantial admixture between populations that can result in spurious assortative or disassortative genetic patterns in pairing (Redden & Allison 2006; Solberg *et al.* 2008; Havlicek & Roberts 2009). These patterns of genetic similarity/dissimilarity between partners can arise because autosomal and MHC genetic variation is structured by ethnicity (Rosenberg *et al.* 2002; Vina *et al.* 2012) where the population frequencies of MHC alleles depend on the geographical location and on the level of population heterogeneity (Prugnolle *et al.* 2005; Solberg *et al.* 2008; Vina *et al.* 2012). A second confounding variable unique to humans is technological alterations in biological phenotypes, which include hygiene and make-up routines, surgery and artificial hormones for controlling contraception (Wedekind *et al.* 1995). However, perfumes appear to be chosen to amplify one's MHC odour profile (Milinski & Wedekind 2001). Lastly, it may sometimes be difficult to standardize experimental designs between studies when using human subjects and these methodological differences can also confound results (Havlicek & Roberts 2009). Several recent reviews have discussed these issues in attempts to reconcile significant and nonsignificant results from over three decades of human MHC mate choice research (Havlicek & Roberts 2009; Winternitz & Abbate 2015). However, only quantitative assessment that explores the magnitude and precision of effects can reveal the ultimate biological importance of phenomena (Nakagawa & Cuthill 2007). Thus, there is a need for a quantitative comparison of human MHC studies. For evolutionary context, nonhuman primates should be relatively free from the confounding biological and technological aspects of human studies and thus should more closely resemble human evolutionary origins than contemporary human populations.

To determine the biological importance of MHC-based mating across human populations and uncover drivers of mate selection, we performed a phylogenetically controlled meta-analysis of published studies. We tested for biologically significant mate choice for MHC dissimilarity and MHC diversity separately. We aimed to identify consistent relationships between MHC and mating across primates, including nonhuman primates, to put human results in context. In addition, we tried to disentangle potential biological or methodological sources of variation in the observed effect sizes to better understand differences between past studies and minimize differences for future studies.

## Materials and methods

### *Literature search*

Data set compilation methods are described in detail in Winternitz & Abbate (2015). Briefly, studies were compiled from the reviews by Havlicek & Roberts (2009) and Setchell & Huchard (2010), from the meta-analysis by Kamiya *et al.* (2014), and additional studies were identified up from 2009 to January 2015 via Web of Science using the topic 'MHC' and 'Major Histocompatibility Complex' and 'mate choice' or 'mate selection' or 'mate preference' and searching within results for 'human' and 'primate'. Studies were listed as testing for human or primate preferences, and for preferences for MHC dissimilarity or diversity/heterozygosity. Studies were included if MHC genotypes [or their approximations via single nucleotide polymorphisms (SNPs), e.g. HapMap data] were obtained for the individuals tested. Studies were excluded if we could not extract the full set of effect sizes that related to the question of MHC influence on mating preferences [e.g. Giphart & D'Amaro (1983) did not provide test statistics for pairwise tests]. We only considered classical MHC class I and class II genes since nonclassical MHC genes, although also involved in immunity, usually have tissue-specific expression and much less diversity indicating different selection schemes (reviewed in Rodgers & Cook 2005). For example, Khankhanian *et al.* (2010) provided data for nonclassical HLA-E, which we did not use. Similarly, Laurent *et al.* (2012) only presented data for nonclassical HLA-L and HLA-J genes. Lists of full references, including references we could not use and explanations for exclusion, are provided in the supplementary materials.

### *Data extraction and effect size calculation*

We chose $r$ effect size (correlation coefficients) as the measure of the association between MHC target (dissimilarity or diversity) and the strength of mating preference/outcome. Studies had mostly measured dissimilarity as categories of allele sharing (e.g. none

© 2016 John Wiley & Sons Ltd

and = >1) and occasionally as allele sequence divergence. Diversity was mostly measured categorically (i.e. homozygous at one or more loci vs. heterozygous) and occasionally continuously (mean heterozygosity over all loci considered). Test statistics other than measures of correlation were converted to *r* effect sizes following (Nakagawa & Cuthill 2007); while there are limitations to converting *r* from summary statistics other than bivariate correlations (Aloe 2015), we stress that these limitations will introduce noise and not bias. Results from Yang *et al.* (2014) could not be directly converted to effect sizes so we calculated values for the data points (points were extracted from Fig. 5 using software DATATHIEF (Tummers 2006)) and fit correlation models to obtain effect size estimates in the desired scale. When studies provided multiple effect sizes that we could not independently evaluate with moderator variables (e.g. results from multiple loci or MHC allele and supertype data), we calculated weighted means by first converting measures to *r* and then weighting them by the underlying sample sizes. We accounted for nonindependence of multiple effect sizes extracted from the same study that remain after calculating weighted means by accounting for the specific hierarchical structure of the data in the appropriate statistical models (see the structure of the mixed model in the next section). For studies that listed effect sizes for MHC similarity preferences, we reversed the sign (i.e. resulting in a negative association for dissimilarity) to align data for preference for dissimilar mates according to the predictions of our focal biological hypothesis. The number of raters was recorded to test for potential effects of sample size on the resulting effect size. The number of individuals rated (number of independent repeats) in the study was recorded to calculate the variance in effect size (variance = $1/(N_{\text{study rated}} - 3)$). When weighted effect size means were calculated, we also recorded the mean number of individuals rated and used this estimate to calculate the variance of the weighted mean. Raw data and converted effect sizes were checked by independent extraction (JA), and any inconsistency was discussed until a consensus was reached (between JW, JA and LG). We converted effect sizes into Fisher's Z (Zr) to stabilize variance across effect sizes, and Zr and its variance (defined above) were used for meta-analyses. The full data set, comprising 58 effect sizes from 30 studies across seven species, and effect size extractions and conversions are provided in the electronic supplementary material. The data set was split to test for MHC dissimilarity and diversity-mediated mating preferences across primates ($N = 41$ and 17, respectively). Human studies greatly outnumber the other species, which may lead to biases in the observed effect size patterns. Therefore, we also analyse them separately from nonhuman primates.

Biological and methodological differences between studies have been shown to predict variation in MHC-mediated mating patterns in human and nonhuman populations (Havlicek & Roberts 2009; Setchell & Huchard 2010; Kamiya *et al.* 2014). We accounted for these potential sources of heterogeneity by considering various moderator variables for the partition of the between-study variance in effect sizes. The following data were extracted from each study as methodological predictors: (i) study ID and (ii) year of publication for publication bias testing, (iii) choice cue used for mating preference (i.e. facial attractiveness, odour attractiveness or mate choice outcome), (iv) the number of individual raters (*N* of rater), (v) multilocus or single locus, (vi) level of population heterogeneity (dichotomously classified as ethnically homogeneous or heterogeneous). This moderator was included to control for artefactual patterns of MHC similarity or dissimilarity that can arise from the pooling of different ethnic groups together in the same sample (Rosenberg *et al.* 1983). Populations were classified as 'ethnically homogeneous' if the study samples fell into single ethnic groups according to the ethnic categories defined by the CIAfactbook (https://www.cia.gov/library/publications/the-world-factbook/index.html), or based on detailed genealogical (i.e. Ober *et al.* 1997) or anthropological records (i.e. Hedrick & Black 1997). All other human populations were classified as 'ethnically heterogeneous'. All nonhuman primate populations were considered 'homogeneous' as they most likely represented single isolated populations (Setchell & Huchard 2010). (vii) Contraceptive pill use can potentially reverse previous preferences (Wedekind *et al.* 1995) so we ran all models excluding pill-use effect sizes ($N = 4$), but results remained essentially unchanged when all effect sizes were used (both sets of results provided in all tables). We included pill use as a moderator for human odour preference studies (female-pill users, female nonpill users). Biological predictors included (viii) species, (ix) choosy sex (i.e. the unit of investigation: males, females or pairs), (x) MHC class (class I, class II or both) and (xi) relative testes size as a proxy for mating system (Harcourt *et al.* 1981, 1995; Dixson & Anderson 2004; Figure S1, Supporting information). Mating system strongly impacts individual mating strategies and population genetic structure (Sugg *et al.* 1996) and so could influence the expression of MHC-mediated mate choice (Setchell & Huchard 2010).

Unfortunately, background genetic measures of dissimilarity and diversity were not available for the majority of studies in our data set, and thus, we could not include this potentially important moderator in our

© 2016 John Wiley & Sons Ltd

analyses. However, we were able to extract a limited number of effect sizes based on neutral markers, which were found to show positive but nonsignificant correlations with MHC-based effect sizes (MHC dissimilarity correlation (95% highest posterior density, HPD) = 0.164 (−0.421 to 0.747), $N = 11$; MHC diversity correlation (95% HPD) = 0.110 (−0.708 to 0.942), $N = 5$, Figure S11, Supporting information). See Supplementary Text for analysis details). This suggests that MHC-mediated effects were largely independent of genome-wide effects.

### *Statistical analyses*

*Meta-analytic procedures.* There were three causes of nonindependence in our data sets: (i) more than one effect size was extracted from a study, (ii) multiple effect sizes were available for the same species, and (iii) species share evolutionary history making effect sizes confounded by the phylogeny of species. We used phylogenetic mixed-effects modelling that includes random effects to account for nonindependence caused by study-, species- and phylogeny-specific effects (Hadfield 2010; Hadfield & Nakagawa 2010; Nakagawa & Santos 2012). A phylogenetic tree was obtained by trimming the primate 10KTree (Arnold *et al.* 2010) using the *drop.tip* function from the APE package version 3.4 (Paradis *et al.* 2004). The deviance information criterion (DIC) was computed for all models considering different random-effect structure (Table 1), and top model selection was based on DIC values, where the lowest DIC is considered the best, models within 2 DIC units are considered equivalent, and a change in DIC of 4 or more significantly improves prediction (Spiegelhalter *et al.* 2002). We calculated phylogenetic heritability or phylogenetic signal, $H^2$, as the proportion of total variance in Zr that can be explained by phylogenetic variance (Hadfield & Nakagawa 2010; de Villemereuil & Nakagawa 2014), equivalent to Pagel's λ (Pagel 1999). Our data sets of MHC dissimilarity and MHC diversity were limited (dissimilarity $N = 41$; diversity $N = 17$) and were comprised of 28 and 11 studies, respectively, and seven species each. Preliminary exploratory analysis of the data revealed that the factors study ID and species were strongly associated, with each study focusing on a single species. We compared models for all combinations of the random factors: study ID, species and phylogeny. We found that including study ID greatly improved model fit, but models adding species, phylogeny or both were all within 1 deviance information criterion (DIC) from the top model (Table 1) and essentially equivalent in terms of prediction (Spiegelhalter *et al.* 2002). Therefore, to avoid potentially overfitting the models, we chose to include only study ID and phylogeny to control for phylogenetic pseudoreplication in multispecies models, and to include study ID in human-only models.

Meta-analyses were conducted with generalized linear mixed-effects models with Markov chain Monte Carlo techniques using the R package MCMCGLMM (Hadfield 2010). We present details on MCMCGLMM model specification and diagnostics in the supplementary material. Briefly, all models were fit using an uninformative inverse gamma prior for all random effects and residuals and were checked for sensitivity to prior specification and convergence across independent model runs (following Wilson *et al.* 2010). Each model was run for 3 million iterations, sampling every 500 after discarding one-million, and this process was repeated for each model to confirm stability of results. We first ran intercept-only mixed models (with random effects) to determine the mean effect size across all studies and for humans and nonhuman primates separately. We tested whether specifying priors based on the effect size estimates from mammalian MHC mate choice (Kamiya *et al.* 2014) would improve model fit and reduce variance around posterior estimates (Garamszegi *et al.* 2009), but results were essentially identical to those obtained with uninformative priors (Tables S1, S2, S6, Supporting information).

*Heterogeneity estimation.* Variation in observed effect sizes between studies is composed of both real differences in mating outcome (effect size heterogeneity) as well as random error. To estimate effect size heterogeneity, we used $I^2$ (Higgins & Thompson 2002; Higgins *et al.* 2003) modified for multilevel meta-analytic models (Nakagawa & Santos 2012). Low, moderate and high heterogeneities refer to $I^2$ of 25%, 50% and 75%, respectively (Higgins *et al.* 2003). The intercept-only models indicated high heterogeneity (>75%) in effect sizes, and while this was mostly explained by the random effects of study ID and phylogeny, substantial residual heterogeneity persisted (26% for dissimilarity, 14% for diversity, Table 2).

We next constructed a series of meta-regression models to identify the most important moderators (listed above) that explained substantial residual heterogeneity in effect sizes (Nakagawa & Santos 2012). We conducted univariate fixed-effect mixed models to estimate the mean effect size for each moderator separately (we avoided complex models with multiple predictors given the limited sample size). Models with categorical moderators were run without the intercept to test each trait against no effect. Parameter estimates were based on posterior means and estimates with highest posterior density (HPD) intervals that do not cross zero are inferred to represent real effects. All effect sizes are

© 2016 John Wiley & Sons Ltd

**Table 1** Comparison of models with varying random effects and their deviance information criteria (DIC) values for effects of MHC dissimilarity and diversity on mate choice for all species, only humans, and only nonhuman primates. All models include the intercept and the listed random effects. Values in bold specify the random effects used in subsequent models. Models within 2 DIC units are essentially equivalent (Spiegelhalter *et al.* 2002), so we chose to include random effects that would account for study and phylogenetic nonindependence without overfitting the models

| | | All | | Humans | | Nonhuman primates | |
|---|---|---|---|---|---|---|---|
| Model ID | Random effects | DIC | ΔDIC | DIC | ΔDIC | DIC | ΔDIC |
| MHC dissimilarity | | | | | | | |
| Model0 | No random | −8.76 | 43.00 | −4.20 | 34.53 | −14.76 | 0.54 |
| **Model1** | ID | −51.09 | 0.68 | **−38.73** | **0.00** | −15.30 | 0.00 |
| Model2 | Species | −8.26 | 43.51 | . | . | −15.30 | 0.01 |
| Model3 | Phylogeny | −8.57 | 43.20 | . | . | −15.13 | 0.17 |
| **Model4** | ID + Phylogeny | **−51.77** | **0.00** | . | . | **−15.28** | **0.02** |
| Model5 | ID + Species | −51.72 | 0.05 | . | . | −15.26 | 0.04 |
| Model6 | ID + Species + Phylogeny | −51.69 | 0.08 | . | . | −15.16 | 0.14 |
| MHC diversity | | | | | | | |
| Model0 | No random | −45.12 | 1.26 | −25.05 | 0.00 | −19.85 | 0.00 |
| **Model1** | ID | −45.73 | 0.65 | −24.82 | **0.23** | −19.32 | 0.54 |
| Model2 | Species | −46.06 | 0.32 | . | . | −19.37 | 0.48 |
| Model3 | Phylogeny | −46.38 | 0.00 | . | . | −19.61 | 0.24 |
| **Model4** | ID + Phylogeny | **−46.31** | **0.07** | . | . | **−19.33** | **0.53** |
| Model5 | ID + Species | −46.11 | 0.27 | . | . | −19.24 | 0.61 |
| Model6 | ID + Species + Phylogeny | −46.17 | 0.21 | . | . | −19.35 | 0.51 |

reported as Fisher's normalized correlation coefficients (Zr) with 95% confidence intervals. In ecological literature, $r \approx 0.1$ (Zr $\approx$ 0.10) is generally considered a small effect, $r \approx 0.3$ (Zr $\approx$ 0.31) a medium effect and $r \approx 0.5$ (Zr $\approx$ 0.55) a strong effect (Cohen 1988; Møller & Jennions 2002).

*Publication bias and power analysis.* We tested for publication bias using four different approaches given that they have different advantages and disadvantages. First, we applied Egger's regression (Egger *et al.* 1997) on meta-analytic residuals instead of effect sizes, which can better distinguish between publication bias and other sources of heterogeneity (Egger *et al.* 1997; Sutton *et al.* 2011; Kamiya *et al.* 2014). If the regression of the standard normal deviate (residuals divided by the standard error) on precision has an intercept different from zero at 90% confidence, then there is evidence of bias favouring publication of less precise yet significant results (Egger *et al.* 1997). Second, we tested for temporal bias in publication results (e.g. if nonsignificant studies are suppressed immediately after the first significant publication) by including the publication year of the study as a moderator in the meta-analytic model. We also used Spearman's rank to test for significant correlations between effect size and year of publication. Third, to assess the impact of publication bias and test the robustness of our results, we used the nonparametric trim and fill method (Duval & Tweedie 2000a,b) in the METAFOR R package (Viechtbauer 2010). This method adjusts the mixed-model intercept for potentially missing studies and the difference is added to the original meta-analysis model intercept (and credible interval; following Sutton *et al.* 2011). Fourth and finally, as bias for publications with significant results can rely more on the *P*-value than on the effect size, we used the *P*-curve method to test whether the distribution of significant *P*-values, the '*P*-curve', indicates that our studies have evidential value and are free from 'p-hacking' (Simonsohn *et al.* 2014a, b). While problems in identifying publication bias using the *P*-curve method have been identified (Bishop & Thompson 2016; Bruns & Ioannidis 2016), we controlled for false-negative results by ensuring that all data entered into analysis met the three required assumptions set by Simonsohn *et al.* (2014b). Specifically, these assumptions are that *P*-values are (i) associated with the hypothesis of interest, (ii) statistically independent from other selected *P*-values, and (iii) distributed uniformly under the null hypothesis of no bias.

We tested the robustness of our results by conducting retrospective power analyses to evaluate whether our sample size (number of effect sizes) was sufficient to have a high chance of detecting a biologically significant effect. We used a prespecified effect size of 0.15 (explaining 2.2% of the variation in mating patterns) which fell within the observed range of effect size estimates from the meta-analysis of Kamiya *et al.* (2014)

© 2016 John Wiley & Sons Ltd

**Table 2** Heterogeneity estimates and deviance information criteria (DIC) for a set of random-effect-only meta-analytical models for MHC dissimilarity and diversity. The heterogeneity ($I^2$) value is the percentage variance from a particular random factor over the sum of all variance components plus the mean variance and was calculated from posterior means. The total $I^2$ (and the 95% HPD) is the sum of all variance components. The mode total variance is shown for comparison (where similar values indicate stable models). Phylogenetic heritability ($H^2$) is the proportion of variance that can be explained by phylogenetic variance (Hadfield & Nakagawa 2010). The final random-effect models used in subsequent analyses are highlighted in bold. Rationale for using the bold models is explained in the methods. Model ID refers to models from Table 1

| Data set | *N* | Model ID | DIC | Study $I^2$ | Residual $I^2$ | Phylogeny $I^2$ | Mean total $I^2$ (HPD) | Mode total $I^2$ | Mean $H^2$ (HPD) | Mode $H^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Dissimilarity | | | | | | | | | | |
| All | 41 | Model0 | −8.77 | . | 85.49 | . | 85.49 (77.66–92.73) | 86.74 | . | . |
| | | **Model4** | **−51.77** | **29.20** | **32.09** | **27.02** | **88.32 (78.90–97.50)** | **88.80** | **0.30 (0.01–0.78)** | **0.04** |
| All | 37 | Model0 | −7.88 | . | 86.88 | . | 86.88 (79.17–93.52) | 88.53 | . | . |
| (no pill) | | **Model4** | **−60.99** | **37.01** | **25.72** | **26.57** | **89.30 (80.63–98.11)** | **88.68** | **0.29 (0.004–0.77)** | **0.04** |
| Human | 35 | Model0 | −4.20 | . | 87.69 | . | 87.69 (80.17–94.37) | 89.51 | . | . |
| | | **Model1** | −38.73 | **39.07** | **47.13** | . | **86.21 (76.84–93.97)** | **87.72** | . | . |
| Human | 31 | Model0 | −3.56 | . | 89.01 | . | 89.01 (81.73–95.08) | 89.81 | . | . |
| (no pill) | | **Model1** | **−47.08** | **51.54** | **36.31** | . | **87.85 (78.79–94.73)** | **89.4** | . | . |
| Nonhuman | 6 | Model0 | −14.76 | . | 34.14 | . | 34.14 (2.24–76.71) | 9.03 | . | . |
| primate | | **Model4** | −15.28 | **19.64** | **20.45** | **29.67** | **69.75 (36.62–98.89)** | **83.21** | **0.41 (0.003–0.92)** | **0.09** |
| Diversity | | | | | | | | | | |
| All | 17 | Model0 | −45.13 | . | 26.71 | . | 26.71 (2.97–60.73) | 11.53 | . | . |
| | | **Model4** | −46.31 | **15.71** | **13.76** | **34.38** | **63.84 (30.49–94.67)** | **63.68** | **0.50 (0.04–0.96)** | **0.16** |
| All | 16 | Model0 | −41.30 | . | 28.54 | . | 28.54 (2.15–62.28) | 9.09 | . | . |
| (no pill) | | **Model4** | **−42.67** | **16.73** | **14.53** | **32.95** | **64.21 (33.90–96.71)** | **77.24** | **0.48 (0.03–0.94)** | **0.17** |
| Human | 10 | Model0 | −25.05 | . | 27.38 | . | 27.38 (1.96–65.87) | 8.50 | . | . |
| | | **Model1** | −24.82 | **22.93** | **22.19** | . | **45.12 (10.04–81.55)** | **42.10** | . | . |
| Human | 9 | Model0 | −21.34 | . | 30.33 | . | 30.33 (1.65–68.71) | 7.18 | . | . |
| (no pill) | | **Model1** | **−21.34** | **24.39** | **23.37** | . | **47.75 (12.78–84.87)** | **33.18** | . | . |
| Nonhuman | 7 | Model0 | −19.85 | . | 38.41 | . | 38.41 (3.64–84.60) | 10.62 | . | . |
| primate | | **Model4** | −19.33 | **25.80** | **17.59** | **35.71** | **79.09 (47.31–99.44)** | **93.10** | **0.44 (0.004–0.94)** | **0.03** |

that investigated MHC-mating patterns across vertebrates [dissimilarity Zr (HPD) = 0.064 (−0.080 to 0.193); diversity Zr = 0.113 (−0.004 to 0.237)]. We used this prespecified effect size to represent the minimal biologically significant effect, and we used the observed mean variance of the effect sizes following recommendations of Thomas (1997). We conducted our power analyses for meta-analytic random-effects models for low, medium and high levels of heterogeneity using the methods of Hedges & Pigott (2001).

All statistical analyses except for the *P*-curve method (implemented at http://www.p-curve.com/) were carried out in the R environment (version 3.2.1; R Core Team 2015), and all R code is provided in the online appendix. The R packages we used were APE version 3.5 (Paradis *et al.* 2004), MCMCGLMM (Hadfield 2010), METAFOR (Viechtbauer 2010), PHYTOOLS (Revell 2012), PLOTMCMC (Magnusson & Stewart 2014) and SHAPE (Soetaert 2014).

## Results

The main results are presented in the next four sections, and detailed results of our meta-regression analysis can be found in the supplementary materials (Tables S1–S17, Figures S1–S11, Supporting information).

### *Preference for MHC dissimilarity*

The mean effect size calculated over all studies (excluding contraceptive pill users) indicated no significant correlation between MHC dissimilarity and mating outcome (intercept-only posterior mean Zr (95% HPD) = 0.044 (−0.174 to 0.289), $N = 37$). The total heterogeneity (total $I^2$) in effect sizes was large (89%) and could mostly be explained by the two random factors ($I^2_{ID} = 37\%$, $I^2_{phylogeny} = 27\%$), with substantial residual variance remaining ($I^2_{residual} = 26\%$), shown in Table 2. We ran univariate models to identify moderators potentially explaining residual heterogeneity, and while no moderator was identified as significant (Table S1, Supporting information), the moderator 'choosy sex' showed significant contrasts between the categories of males and pairs (contrast $P = 0.02$, Fig. 1a, Table S11, Supporting information). In other words, studies using mated pairs had greater effect sizes for MHC-similar mates than studies investigating male

© 2016 John Wiley & Sons Ltd

preferences. This effect was driven by human studies, which had significant contrasts not found for nonhuman primates (human contrast $P = 0.031$, Table S11, Supporting information). Phylogenetic heritability in MHC-dissimilar mating patterns was low (mean $H^2 = 0.29$ (0.004–0.77), mode $H^2 = 0.04$, Fig. 2a, Table 2), and we note that random effects are bound to be positive and their posterior distributions will never overlap zero (Wilson *et al.* 2010). Thus, the meaningfulness of the random effect of phylogeny cannot be based on its nonzero posterior distribution. Despite its effect of increasing the HPD intervals for model estimates, we retained phylogeny as a random effect to control for pseudoreplication in multispecies models. We investigated the impact of phylogeny on the meta-mean effect size of MHC dissimilarity by comparing model DICs and posterior estimates for models with and without phylogeny, but results were qualitatively similar and nonsignificant (Table S13, Supporting information).

When human and nonhuman primate studies were examined separately, neither had significant associations between MHC dissimilarity and mating outcome (human Zr = −0.022 (−0.107 to 0.073), $N = 31$, nonhuman primate Zr = 0.109 (−0.194 to 0.404), $N = 6$). Mean effect sizes calculated for these two taxonomic groups were not statistically differentiable. In intercept-only models, total heterogeneity was high for both humans and nonhuman primates ($I^2_{total}$ = 88% and 70%, respectively) with substantial residual variance in humans specifically ($I^2_{residual}$ = 36% vs. 20% in primates, Table 2). Moderators tested separately in univariate models for humans, in models by human choice cue and in models for primates were all nonsignificant and did not reduce residual heterogeneity (Tables S2–S4, S9,

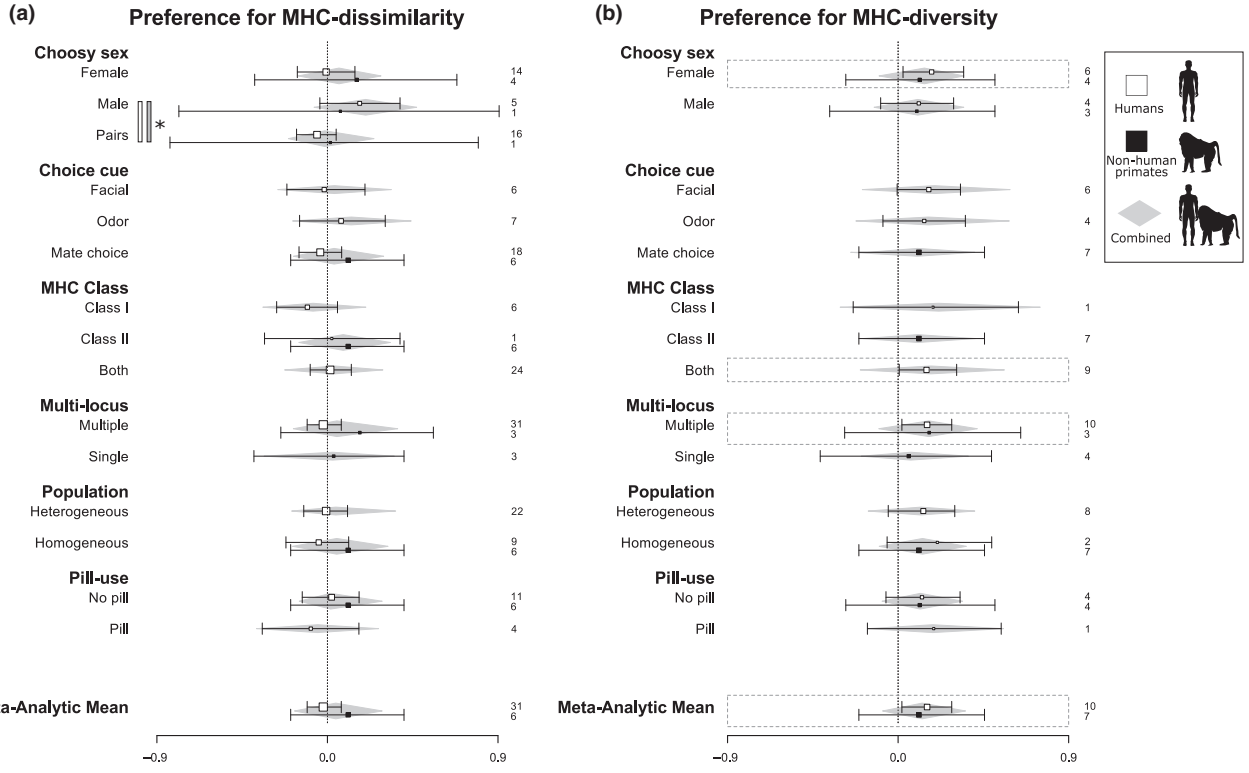


**Fig. 1** Effect of moderators on strength of MHC-associated mating outcome. This figure is a forest plot of effect size for categorical moderators for (a) preference for MHC dissimilarity and (b) preference for MHC diversity. Positive posterior estimates of Fisher's Z transform of the correlation coefficient (Zr) indicate positive associations, whereas negative estimates indicate negative associations between MHC target and mating outcome. Boxes show the mean posterior estimate from the model, and error bars represent the 95% highest posterior density (HPD) interval. Numbers on the right-hand side of each panel indicate the number of effect sizes in each subgroup. White boxes indicate model estimates from human data, black boxes are from nonhuman primate data, and grey polygons indicate estimates from models using all data. We removed pill-user effect sizes from the combined primate and human data set to control for its potentially confounding effect (see Methods) and tested pill use as a moderator for studies that investigated its affect or controlled for it. The meta-analytic mean is from the intercept-only model run with study ID and phylogeny as random effects (and study ID as the random effect for the human data model). Dashed boxes highlight mean estimates with HPD intervals that do not overlap zero. Vertical bars indicate significant contrasts between moderator categories for humans and all data (*$P$-value < 0.05, see Table S10, Supporting information).

© 2016 John Wiley & Sons Ltd

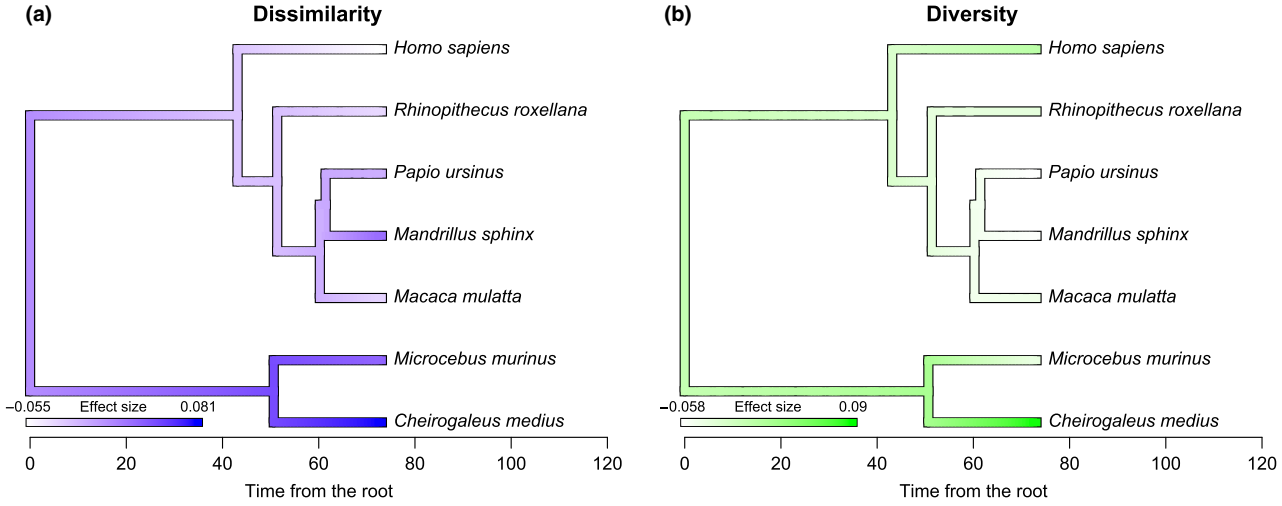


**Fig. 2** Phylogenetic heritability of effect sizes for MHC-dissimilar and MHC-diverse mating patterns. We calculated the residuals of Zr effect sizes due to phylogenetic relatedness and computed the mean for each species to graphically represent the heritability of effect sizes for (a) MHC-dissimilar mating patterns (heritability posterior mean (HPD) = 0.30 (0.01–0.78)) and (b) MHC-diverse mating patterns (heritability posterior mean (HPD) = 0.50 (0.04–0.96)). Colours indicate the strength of phylogenetic signal.

Supporting information, Figs 1a and 3), but the direction of effect sizes for primates was consistent for dissimilarity (Figs 1a and 2a). Phylogenetic heritability in MHC-dissimilar mating patterns among nonhuman primates was low but present (mean $H^2$ = 0.41 (0.003–0.92), mode $H^2$ = 0.09, Table 2).

### *Preference for MHC similarity*

As a *post hoc* investigation to explain the large residual heterogeneity in effect sizes among the human dissimilarity data set ($I^2_{\text{residual}}$ = 36%, Fig. 4a), we specifically tested for the effects of population ethnic heterogeneity on patterns of MHC-assortative mating among experimental and observational studies (predicted by Rosenberg *et al.* 1983 and Havlicek & Roberts 2009). We ran mixed-effects models for the subset of dissimilarity effect sizes that were negative (indicating preferences for similarity). We found that preference for MHC similarity was significant in ethnically heterogeneous population samples, but not among those that were homogeneous. This result was found only in observational (mate choice) studies, and not among those using experimental approaches (odour and facial preference combined; Fig. 5, Table S5, Supporting information).

### *Preference for MHC diversity*

For diversity models, we present results for effect sizes ($N$ = 17) including one from contraceptive pill-using women because pill use was never shown to have an effect on preferences for MHC-diverse mates, its exclusion did not reduce heterogeneity, and model results are virtually identical whether this effect size was excluded or not (both results presented in all tables for comparison). Across all effect sizes combining humans and primates, there was a nonsignificant trend for MHC diversity to be positively associated with mating outcome (posterior mean Zr (HPD) = 0.128 (−0.064 to 0.373), $N$ = 17; Fig. 4b shows significant raw mean Zr effect sizes and results from random-effect models without accounting for study ID and phylogeny). Total heterogeneity was large ($I^2_{\text{total}}$ = 64%) but was mostly accounted for by differences between studies and phylogeny ($I^2_{\text{ID}}$ = 16%, $I^2_{\text{phylogeny}}$ = 34%, Table 2). Residual heterogeneity was low ($I^2_{\text{residual}}$ = 14%) and could only slightly be reduced by the addition of the moderator relative testes size as a fixed effect (Table S10, Supporting information). Univariate models of moderator effects were nonsignificant (Table S6, Supporting information). The model including relative testes size was also nonsignificant, but the negative association between relative testes size and preferences for MHC diversity was trending towards biological significance (intercept (HPD) = 0.246 (−0.019 to 0.581); posterior mean testes (HPD) = −0.218 (−0.479 to 0.047), Table S6, Supporting information). Phylogenetic heritability in strength of MHC diversity mating patterns was moderate (mean $H^2$ = 0.50 (0.04–0.96), mode $H^2$ = 0.16, Table 2, Fig. 2b).

Examined separately, humans showed a significant association for more MHC-diverse individuals to be preferred as mates (humans = 0.153 (0.020–0.283), $N$ = 10, Fig. 1b) while nonhuman primates had a nonsignificant trend for mate choice for diversity (primates = 0.110 (−0.207 to 0.456), $N$ = 7, Fig. 1b). Total

© 2016 John Wiley & Sons Ltd

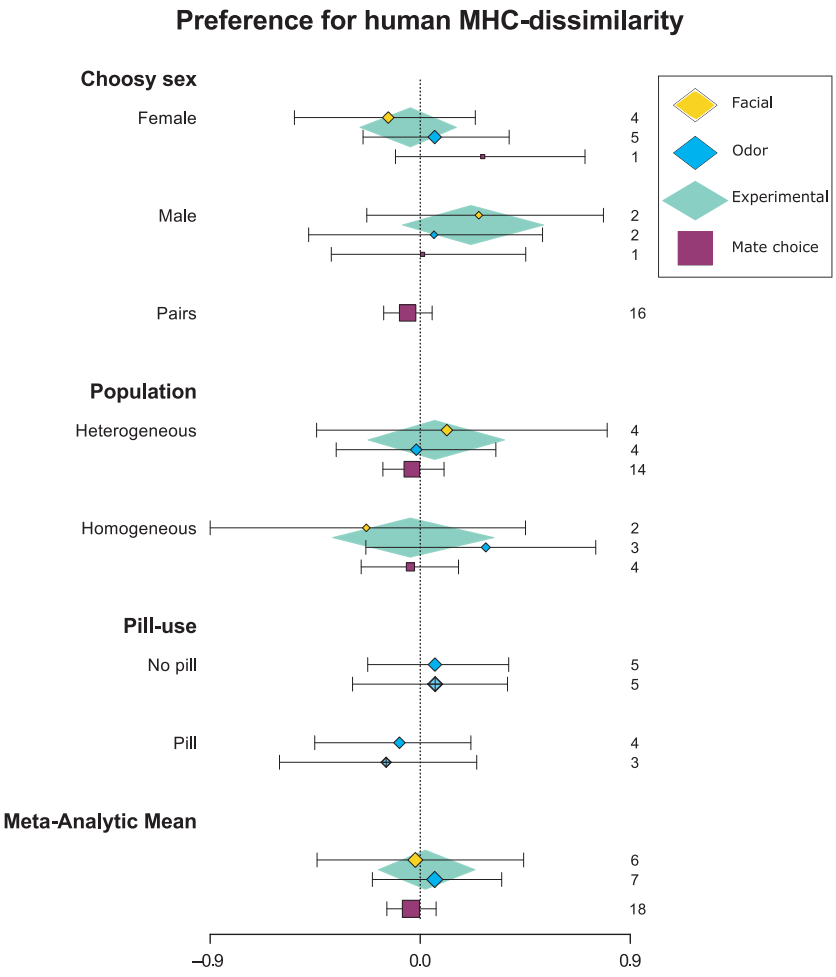


**Fig. 3** Effect of moderators on strength of human MHC dissimilarity mating preferences (experimental designs) and mate choice (observational data). Zr effect sizes for human data were investigated by choice cue category, including facial preferences (white/yellow diamond), odour preferences (dark grey/blue diamond), these experimental categories combined (light grey/green polygon), and mate choice studies (black/maroon square). Points show the mean posterior estimate from the model, and error bars represent the 95% HPD interval. Numbers on the right-hand side of each panel indicate the number of effect sizes in each subgroup. We ran all models excluding potentially confounding pill-use effect sizes, except those models specifically testing for pill-use effect (only available for odour preference studies). The cross-diamond model estimates include only studies that had dichotomously classified raters as pill users or not [i.e. Santos *et al.* (2005) provided an effect size for a group of raters in which 30% were taking birth control pills]. The meta-analytic mean is from the intercept-only model run with study ID as a random effect. [Colour figure can be viewed at wileyonlinelibrary.com]

heterogeneity was moderate for humans ($I^2_{total}$ = 45%) and high for nonhuman primates ($I^2_{total}$ = 79%), and mostly explained for primates by the random effects of study ID and phylogeny ($I^2_{ID}$ = 26%, $I^2_{phylogeny}$ = 36%, Table 2). Residual variance was not substantially reduced after the addition of moderators as fixed effects for humans (Table S10, Supporting information) and these mixed models were essentially equivalent to the intercept-only model (all ΔDIC < 2). The addition of the moderator relative testes size slightly reduced residual heterogeneity and DIC for nonhuman primates compared with the intercept-only model (Table S10, Supporting information), but not substantially (ΔDIC < 2). When examining categorical level differences in MHC diversity effect sizes using univariate models, we found no significant moderator for nonhuman primates (Table S8, Supporting information). In contrast, humans showed stronger preferences for MHC-diverse mates for the categories of choosy sex (female rater) and MHC class (when both classes were investigated together, using multiple loci; Table S7, Supporting information, Fig. 1b).

Raw Zr effect sizes for all primates were positively associated with preferences for MHC diversity and showed significant means when not accounting for study ID and phylogenetic pseudoreplication (Fig. 4b). All our mixed models including phylogeny as a random effect had wide 95% confidence intervals for the phylogenetic heritability of effect sizes (proportion of total variance explained by phylogenetic variance, see Table 2). Thus, to test the stability of our results, we reran mixed models for combined data sets and for nonhuman primate data sets with the alternative random-effects ID + species, and ID + species + phylogeny and found qualitatively similar results (Table S13, Supporting information), indicating our conclusions are robust to the random-effect structure employed.

### *Publication bias and power analysis*

We found no evidence for publication bias in the data sets. Egger's Regression tests indicated the intercepts for MHC dissimilarity and diversity data sets were not significant at 90% confidence intervals (Table S14,

© 2016 John Wiley & Sons Ltd

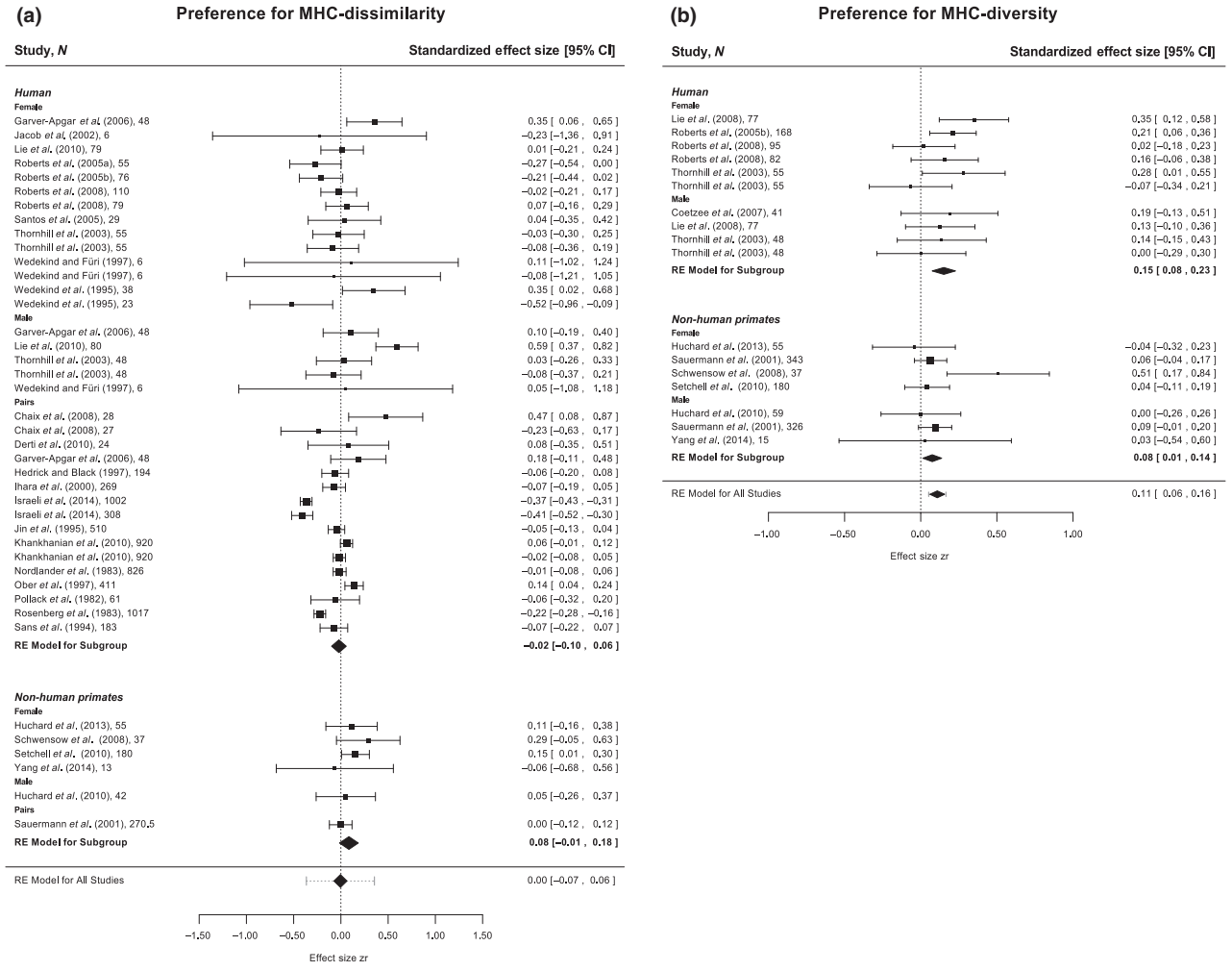


**Fig. 4** Forest plot of original Fisher's Z-transformed (Zr) effect sizes extracted from each study. Zr effect sizes and associated variance were extracted from each study for (a) preference for MHC dissimilarity (positive Zr values indicate preference for dissimilarity, negative for similarity) and (b) preference for MHC diversity. Black diamonds indicate means from random-effect (RE) models for humans and nonhuman primates presented separately, and for all data combined, and error bars represent the meta-analytic variance ($1/(N_{\text{study}} - 3)$). *N* is the number of independent targets per study.

Figure S2 and S3, Supporting information). The slope for MHC dissimilarity was significantly negative, indicating that studies with larger sample sizes (and less variance) showed stronger preference for MHC similarity (Table S14, Figure S2a, Supporting information). This effect was largely due to human correlative studies that had large sample sizes but did not control for ethnicity and assortative mating biases. Trim and fill analyses on mixed-model residuals (following Nakagawa & Santos 2012) were nonsignificant and suggested there were no missing studies on the left-hand side of the funnel plots (Table S14, Figures S2–S3, Supporting information). The sensitivity adjustment of 0.002 to the original intercept-only mean for human MHC diversity results would increase the significance (adjusted meta-mean (95% HPD) = 0.155 (0.022–0.285)). We found no evidence for temporal bias using year as a moderator, nor using Spearman's rank correlation between effect size and year of publication (all $P > 0.05$, Table S16, Supporting information). Finally, *P*-curve analyses for MHC dissimilarity, similarity and diversity data sets suggest that studies contained significant evidential value and showed no evidence of intense p-hacking (Table S16, Supporting information). However, study sets in general had low likelihoods to detect a true effect (average power = 24% for dissimilarity, 57% for similarity, 33% for diversity).

Power analyses on our meta-analytic results indicated that the combined primate and human data sets for MHC dissimilarity and diversity had high power (77.2%–98.5%) to detect a biologically significant mean effect of 0.15 at an alpha value of 0.05 across low,

© 2016 John Wiley & Sons Ltd

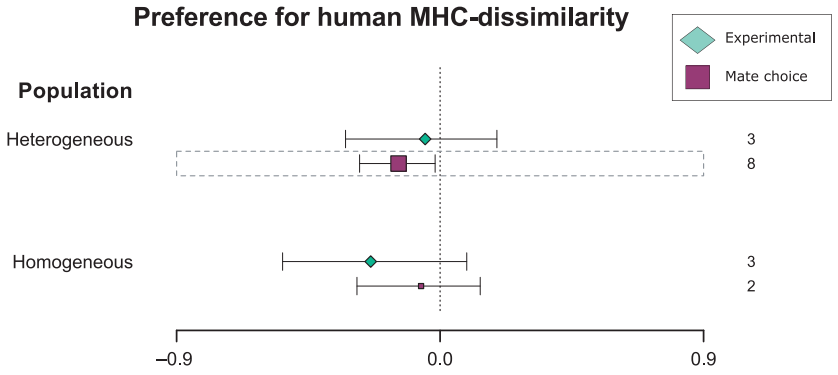


**Fig. 5** Effect of ethnic homogeneity of the population on the strength of human preference for MHC similarity. We investigated the effect of the moderator 'population' on the subset of Zr effect sizes that were negative (indicating preference for similarity). Models with the moderator 'population' were run between choice cue categories for combined experimental studies (odour and facial preference studies, light grey/green diamond) and for mate choice studies (black/maroon square). Points show the mean posterior estimate from the model, and error bars represent the 95% HPD interval. Numbers on the right-hand side of each panel indicate the number of effect sizes in each subgroup. Effect sizes from ethnically heterogeneous populations for mate choice studies (but not experimental studies) had a significant posterior mean estimate for similarity (mean Zr (95% HPD) = −0.142 (−0.275 to −0.017)). [Colour figure can be viewed at wileyonlinelibrary.com]

medium and high heterogeneity values (Table S17, Supporting information). Power was lower for nonhuman primate data sets (dissimilarity = 43.8%; diversity = 58.2% power at high heterogeneity) and a total of 17 and 13 effect sizes, respectively, would be required to reach 80% power for an overall effect size of 0.15 at alpha = 0.05 at high heterogeneity (Table S17, Supporting information). Splitting the human data set by experimental choice cue category, we should have high power to detect a biologically meaningful effect size, if it were present, for mate choice and facial preference sample sizes even at high heterogeneity (98.8% and 64.2% power, respectively). This indicates our non-significant meta-analytic results for human mate choice and facial preference category were likely not due to sample size limitations. In contrast, given the observed variation in effect sizes, we would need 84 odour preference effect sizes to detect biological significance at alpha = 0.05 and high heterogeneity (Table S17, Supporting information).

## Discussion

There is substantial evidence that immunity genes play direct and (less well-supported) indirect roles in mate choice across vertebrates (Kamiya *et al.* 2014). These roles could function to promote diverse immunological repertoires in mates and offspring to respond to diverse parasite attack. Yet nearly four decades of research into this phenomenon in humans has yielded puzzling results. Taking a quantitative meta-analysis approach to put these results in their evolutionary context, we sought to identify patterns of consistency across studies in humans and nonhuman primates. We found a suggestive trend for choice of MHC-diverse mates across primates and clear support in humans, but inconsistency for MHC-dissimilar mating preferences.

### *Humans select more MHC-diverse mates*

Overall, we found a systematic trend for primates and a significant association for humans to prefer more heterozygous mates at MHC sites. In humans, congruent to findings from a range of nonhuman vertebrates (Kamiya *et al.* 2014), we found stronger evidence for female preferences for MHC diversity, and when multiple MHC classes and loci were considered together. This may indicate that females receive greater evolutionary fitness benefits than males from selecting more MHC-diverse mates, and thus, the ability to identify heterozygosity of potential mates is particularly important. Moderate phylogenetic heritability observed in our study implies there may be evolutionary constraints on the expression of mate choice for MHC heterozygous mates. This could be related to olfactory signalling potential that varies among species (Niimura 2009). Our findings also suggest that power to detect significant effects increases when using more MHC classes and loci – allowing for greater variation in individual diversity. Nonhuman primate studies did not use both MHC classes and they rarely used multiple loci in their mate choice research, so the comparatively weaker effect sizes could be due to the lower variation available compared with human studies. Additionally, nonhuman primate effect sizes measured mating outcomes where MHC-mating effects, if existing, are likely to be smaller

© 2016 John Wiley & Sons Ltd

than those from more controlled human experimental studies that measured mating preferences. We did not find mean effect size differences between human odour and facial preference tests, suggesting that humans may be similarly sensitive to both odour and visual cues of diversity (Penn & Potts 1998).

Attractiveness has been shown to correlate with health across cultures (Gangestad & Buss 1993; Hume & Montgomerie 2001; Gray & Boothroyd 2012; Rantala *et al.* 2012) and universal standards of attractiveness based on health and fitness cues (Langlois *et al.* 2000) may be advertising protective immunogenetic diversity (Hamilton & Zuk 1982). Facial attractiveness may be a particularly effective indicator of health and condition, and of individual heterozygosity (Roberts *et al.* 2005b; Lie *et al.* 2008). To illustrate, the faces of mixed-ethnicity individuals, who tend to be more heterozygous than average, have been shown to be more attractive than single-ethnicity faces across genders and cultures (Rhodes *et al.* 2005; Lewis 2010; Little *et al.* 2012). In addition, work using facial image manipulations demonstrated that cues to heterozygosity are themselves attractive independent of other accompanying indicators of heterosis (Lewis 2010; Little *et al.* 2012). How much MHC heterozygosity would make an individual attractive? The majority of human studies (9/10) measured MHC diversity categorically, with individuals homozygous at one or more loci classified as 'homozygous' and all others as 'heterozygous'. Therefore, if increasingly higher levels of heterozygosity or if some optimal level of heterozygosity is favoured remains to be determined. Furthermore, preferences for heterozygosity and dissimilarity need not be exclusive. A preference for both heterozygosity and some degree of similarity is possible (where for each level of similarity, relative heterozygotes are preferred), as demonstrated empirically and by models of correlations between heterozygosity and measures of genetic similarity (Roberts *et al.* 2005b, 2006).

### *Mate choice for MHC dissimilarity is not consistent*

Nonhuman primates tend to consistently prefer MHC-dissimilar mates (Fig. 4a), but our power to detect a significant average effect is limited by our small sample size ($N = 6$) and by the addition of random effects to control for pseudoreplication. However, meta-analytic power analyses predict that 11 additional effect sizes (total $N = 17$) should be sufficient to detect a biologically relevant effect (Zr = 0.15, explaining 2.25% variance) of MHC dissimilarity on mating patterns at 80% power (Table S17, Supporting information). In contrast to the unidirectional trend for dissimilarity in nonhuman primates (posterior mean Zr (HPD) = 0.109 (−0.194 to 0.404)), humans show great variation in direction and magnitude of effect sizes (Fig. 4a). Directional variation could partly be explained by the unit of investigation; we found pairs had significantly stronger preferences for MHC similarity than males. Human males are thought to be less choosy than females as they tend to invest less in offspring, can reproduce at a faster rate and have higher reproductive variance (Trivers 1972; Puts 2012). Thus, this significant contrast may reflect assortative mating in pairs vs. male indifference. The contrast between pairs and females was not significant, as females had a greater range of preferences.

One interpretation is that pairs represent mate choice *outcomes*, in contrast to individual *preferences* of males and females. Because actual choice of partners is influenced by many socio-economic conditions including ethnicity, nationality, family relatedness, phenotypic similarity and spatial segregation (reviewed in Kalmijn 1998; Bovet *et al.* 2012; Nojo *et al.* 2012) that can result in genetically assortative pairings, the apparent preference for MHC similarity in pairs may be coincidental. In effect, ethnic heterogeneity within a sample will produce patterns of positive assortative mating at the ethnic level, where spouses are more similar at a genomewide level than random pairs of individuals (Chaix *et al.* 2008; Laurent & Chaix 2012). We found strong support for this explanation in our findings of significant effects for similarity within couples sampled from ethnically heterogeneous populations, but not from homogeneous populations or from experimental studies that control for potential ethnic biases (Fig. 5). A way to rule out spurious MHC associations stemming from population stratification – differences in subpopulation background allele frequencies (Cardon & Palmer 2003; Derti *et al.* 2010; Laurent & Chaix 2012) – would be to provide measures of neutral genetic similarity of couples. One study adopting this design was able to determine that MHC dissimilarity in couples of European ancestry was independent of genomewide effects, while MHC similarity observed in couples of Yoruban ancestry was most likely a consequence of kinship-based assortative mating (Chaix *et al.* 2008; Laurent & Chaix 2012). Neutral genetic markers would also allow for testing whether the MHC is specifically targeted or whether there is hitch-hiking during assortative or disassortative mating for other traits (Rosenberg *et al.* 1983; Redden & Allison 2006; Laurent & Chaix 2012).

The variation in human MHC dissimilarity preference but single-direction trend in nonhuman primates suggests that selection pressure for MHC-dissimilar mates may be sensitive to environmental or demographic perturbations primarily affecting humans. In contrast to

© 2016 John Wiley & Sons Ltd

nonhuman primates, humans have undergone a recent population expansion (Kaessmann *et al.* 2001) and show extensive levels of admixture (Lawson *et al.* 2012), while most nonhuman primate populations are more genetically isolated and homogeneous, as are, for example, wild great apes (Prado-Martinez *et al.* 2013). Considering that mate choice relies on interpreting cues about traits relative to their background frequency, large changes in the genetic composition of the mating pool may distort signals. For example, if population-specific targets of optimal offspring diversity exist, mixing those populations would produce conflicting optimal targets.

### *Methodological differences among studies*

Even considering only experimental studies with much greater control over the statistical design, there was still high heterogeneity among human dissimilarity effect size magnitude and direction (Table S3, Supporting information, Fig. 3). It could be argued that differences among studies in methods or statistical design are responsible (Havlicek & Roberts 2009). However, the majority of experimental human studies used the same statistical design of Wedekind *et al.* (1995) which treated the chosen individual as the unit of analysis. Only two odour studies used different designs: one (Jacob *et al.* 2002) repeated the analysis using a within-donor design and found that the analysis 'yields virtually identical results' (McClintock *et al.* 2002). The other (Santos *et al.* 2005) used a chi-squared design, but had also had a percentage of participants on birth control, so this effect size along with other effect sizes of pill users was ultimately removed from the data set and could not influence the results.

Sources of heterogeneity can arise through study design, how the outcome is measured, and through real biological differences between populations. Yet we emphasize that studies that differ in their methodologies can be combined for a meaningful meta-analysis. The meta-analysis of Kamiya *et al.* (2014) is a case in point, which combined studies employing a diversity of methods across a range of vertebrates and found a clear general pattern for individuals to prefer MHC-dissimilar and MHC-diverse mates. In fact, combining research based on different methodologies insures that the variance in effect size patterns reflects this process and not any one methodological artefact (Lajeunesse 2010). We can then explore the factors that contribute to variation in MHC mate choice research to synthesize discordant results. One main source of heterogeneity we detected with statistical support was from combining different ethnic groups in a sample for studies that compared observed pairs to randomly created pairs. Another methodological source of heterogeneity we tested was choice cue because preferences based on different stimuli could show different patterns, but we found no strong evidence of this and other studies have found positive correlation between facial and scent attractiveness (Thornhill & Gangestad 1999; Thornhill *et al.* 2003). Additional methodological differences between human studies and their implications for results have been thoroughly discussed elsewhere (Wedekind *et al.* 2002; Havlicek & Roberts 2009; Derti *et al.* 2010; Winternitz & Abbate 2015). We note that we cannot unanimously differentiate whether the differences between study outcomes are caused by methodology or biology, and it is likely that both types of mechanisms are in effect.

### *Study limitations*

One potential source of type I (false-positive) errors is biases in published effect sizes. Publication bias testing showed no evidence for p-hacking but did reveal that the average power of the studies analysed was low, ranging between 24% and 57% between data sets, indicating that true effects may have gone undetected in those studies. Based on our mean effect sizes for dissimilarity ($Zr = 0.044$) and diversity ($Zr = 0.153$), we recommend study samples sizes of ~4051 and ~260, respectively, to detect true effects in primates 80% of the time. The magnitude of these mean effect sizes is typical for ecological data (Møller & Jennions 2002) but this translates to dissimilarity only explaining approximately 0.2% and diversity 2.3% of the variation in primate mating patterns. Clearly, MHC-mediated mate choice in humans and other primates is just one relatively small consideration of many involved in choosing a mate. Another issue regarding type I error is multiple testing. In our study, we used a large number of predictors, the majority being those from Kamiya *et al.* (2014) to get comparable results across different taxa (mammals, nonhuman primates and humans). We also performed a large number of tests for each MHC target. We appreciate that type I errors may occur when testing for multiple predictors, but highlight that the magnitude of the effects may be suggestive for designing future studies. It is reassuring to see that a significant variable in our study (choosy sex) was also found to be significant in a previous meta-analysis on MHC diversity-based mating in vertebrates (Kamiya *et al.* 2014).

Regarding type II (false-negative) errors, retrospective power analysis of our own results indicates we have sufficient sample sizes to have enough power detect an effect of $Zr = 0.15$ (explaining ~2.2% variation in MHC-mating patterns) for human MHC diversity model and models of all human MHC dissimilarity choice cue categories excluding odour preference studies. Sample sizes for nonhuman primates were too low to have a good

© 2016 John Wiley & Sons Ltd

chance of detecting true effects, but high power could easily be achieved with less than 20 additional effect sizes. Noise in the data may increase the risk of not finding biological effects. Therefore, we have tried to draw attention to models where power was limited and new data would be very helpful (i.e. human odour preference models, nonhuman primate models), to help advance the field of MHC-linked mate choice.

Another challenge linked with interpreting meta-analyses conducted across heterogeneous samples and study designs is the issue of confounding factors. Confounding factors not accounted for in original studies pose a substantial problem for the interpretation of all meta-analytical approaches, as they can increase the risk of missing a true effect. Specifically, in our study, only *four* studies in our data set had (i) controlled for Pill effects, (ii) used subjects of the same ethnicity and (iii) conducted an experimental study to control for confounding factors. Two of these studies (Wedekind *et al.* 1995; Wedekind & Furi 1997) investigated odour preferences and found positive effects of MHC dissimilarity ($r = 0.3347$, $r = 0.11$, respectively). The two other studies (Roberts *et al.* 2005a, b) investigated facial preferences and found negative effects of MHC dissimilarity ($r = -0.2632$, $r = -0.2067$). Therefore, even studies that fulfil the strictest of conditions still find opposing effects of MHC dissimilarity on human mate preferences (which may have a biological explanation if the two modalities work in complementary ways to optimize level of MHC diversity for offspring (Roberts *et al.* 2005a)). Our meta-analysis has shown that various moderators can impact the sizes and directions of results investigating MHC-linked primate mating and therefore can point to gaps in the research field to be addressed by future studies. Unfortunately, given our limited sample size, we were not able to incorporate multiple moderators within the same models. Thus, we cannot rule out a significant effect of MHC dissimilarity on primate mate choice under certain conditions and a definitive answer awaits further studies.

### *Evidence for mate choice for MHC optimality?*

Considering only experimental studies that could control for potential socio-ethnic assortative biases we still found wide variation in effect sizes for MHC dissimilarity. It seems prudent to consider that an alternative biological explanation may be at least partially responsible for the variation. The optimality hypothesis predicts that direction of preference either for MHC dissimilarity or some degree of allele matching may depend on the relative allelic diversity in the pool of potential mates (Aeschlimann *et al.* 2003; Milinski 2006). For example, Aeschlimann *et al.* (2003) showed that sticklebacks preferred dissimilar partners in simulated inbred populations and optimally dissimilar partners in simulated outbred populations. High heterogeneity in MHC-dissimilar mating preferences could reflect differences among individuals attempting to achieve 'optimal dissimilarity' by preferring similar or dissimilar mates depending on ecological and demographic context (Jacob *et al.* 2002; Milinski 2006; Roberts 2009). Thus, MHC-based mate choice may be stronger or easier to detect in settings where there is less population genetic diversity and less heterogeneity in other factors which influence mate choice (Ober *et al.* 1997; Jacob *et al.* 2002; Chaix *et al.* 2008). Considering that the diversity of distinct HLA haplotypes (multilocus set of linked alleles) per population typically ranges from 100s to 1000s (Gragert *et al.* 2013), individuals from isolated populations would most frequently encounter only a fraction of the diversity of haplotypes common in more outbred populations. For example, only 10 haplotypes make up the high-frequency majority for the Hutterite community ($N = 1891$ sampled) where a significant preference for MHC haplotype-dissimilar mating was detected (Ober *et al.* 1997). Matching at six alleles (for six-locus haplotypes) could severely limit the potential antigenic detection range and/or interfere with maternal–foetal interactions (Ober *et al.* 1988, 1997; Lashley *et al.* 2015). The Hutterite results are in line with the hypothesis of mate selection disfavouring extreme MHC similarity (Derti *et al.* 2010), and our results also show that experimental study preferences for MHC-similar individuals had relatively few matching alleles between mates (average of 1–3 alleles). It would be helpful if tests for MHC dissimilarity preferences considered a higher range of potentially matching alleles (e.g. allele matching at 0–6 loci). Odour preference studies in particular could employ synthetic peptides that mimic individual alleles to have greater control of the range of allelic diversity (Milinski *et al.* 2005, 2013).

### *Importance of direct and indirect fitness benefits*

Our finding of greater mean effect size for MHC diversity compared with dissimilarity is in line with evidence from across all vertebrates ($Zr = 0.113$ vs. $Zr = 0.064$, respectively (Kamiya *et al.* 2014)). Detecting MHC diversity in a mate may be easier than detecting dissimilarity, as diversity is expected to correlate positively with the mate's phenotypic condition, including health status (Penn *et al.* 2002) or perception of health status through skin condition (Roberts *et al.* 2005), body mass (Thoß *et al.* 2011) and coloration dependent on infection status (Milinski & Bakker 1990), among others. Dissimilarity, on the other hand, should not be reflected by the mate's phenotype alone, as it depends

© 2016 John Wiley & Sons Ltd

only on the combination of both mates genotypes. Thus, this type of preference would require more sophisticated sensory mechanisms including self-referential capabilities. The evolutionary benefits of human mate choice for MHC diversity may include prolonged parental care and reduced risk of contracting disease for a partner and the offspring (Roberts *et al.* 2005b), in addition to the potential indirect benefits from transmission of advantageous genes to offspring by diverse mates (Brown 1997, 1999; Kempenaers 2007). If preferences for MHC-diverse mates are primarily for direct fitness benefits, then resource-based mating systems where direct benefits (e.g. mate protection, provisioning and paternal care) are important should show stronger effects for mate choice. Alternatively, if diversity preferences are primarily for indirect fitness benefits, then nonresource based mating systems should show stronger effects for mate choice. We found preliminary support favouring direct fitness benefits in our trend of more promiscuous mating systems (with larger relative testes size) correlating with weaker mate choice for MHC diversity in primates (Tables S6 and S10, Supporting information). Another piece of support for direct benefits is that our results show stronger mate choice for MHC diversity in humans, where direct benefits are important, than in nonhuman primates. For instance, the expression of a major direct fitness benefit, paternal care, is intense for humans but rare in nonhuman primates (reviewed in Fernandez-Duque *et al.* 2009). This association suggests that more promiscuous mating systems which tend to provide fewer resource-based benefits to females (Clutton-Brock 1989) have weaker effects of mate choice for MHC diversity.

## Conclusions and suggestions for future research

We found clear support for humans and a trend for nonhuman primates choosing more MHC-diverse mates. In contrast, we found extremely high heterogeneity and no such clear pattern in humans for choice for MHC-dissimilar mates. A key driver of this heterogeneity was whether or not ethnic heterogeneity in studies on couples was controlled. High heterogeneity among nonhuman primate studies still showed a consistent direction for MHC dissimilarity, which could stem from methodological differences but also from socio-ecological differences between populations and species (Setchell & Huchard 2010). In fact, we found preliminary evidence that the expression of MHC-based mate choice could depend on the mating system and the reproductive strategies of individuals within those systems, as well as the phylogenetic history among species.

Results of this study show clear priorities in how to design future human and nonhuman primate studies, including studies that: (i) are large scale (≥200 individual targets) and include power analysis, (ii) focus on individuals from (ethnically) homogeneous populations with limited MHC diversity (Ober *et al.* 1997; Jacob *et al.* 2002; Chaix *et al.* 2008) and test socio-ecologically sensitive hypotheses (e.g. how expression of preferences for diversity/dissimilarity/optimality may vary according to mating system or demography; Setchell & Huchard 2010), (iii) explicitly test the optimality hypothesis (with the prediction of less variance around an optimal parental combination of alleles than combination under random mating; Forsberg *et al.* 2007) and consider experimentally adjusting MHC peptide diversity/overlap (Milinski *et al.* 2013), (iv) use multiple loci, including different MHC classes (I and II; Kamiya *et al.* 2014), (v) control for non-MHC variability–key in determining incidental or adaptive MHC-assortative mating (Jiang *et al.* 2013), and (vi) control for biologically confounding variables (e.g. ovulatory status and contraceptive pill use; Wedekind *et al.* 1995).

Finally, we emphatically call for more nonhuman primate studies to improve understanding of the evolutionary trajectory of human mate choice. We hope that our synthesis highlights the need for additional studies of the selective pressure of MHC genotype on mating decisions and provides direction for future research.

## Acknowledgements

We thank T. Lenz, M. Milinski and the four anonymous reviewers for constructive feedback that substantially improved our final manuscript. Funding support for JW came from the Alexander von Humboldt Foundation, the European Social Fund (ESF) and the state budget of Czech Republic through the Operational Program Education for Competitiveness (OPEC), Reg. No. CZ.1.07/2.3.00/30.0048. Funding support for JA came from the University of Berne. JH is supported by the Czech Science Foundation Grant (14-02290S). LZG was supported by funds from the Ministry of Economy and Competitiveness in Spain (CGL2015-70639-P), and by the National Research, Development and Innovation Office in Hungary (NKFIH, K-115970).

## Conflicts of interest

The authors report no conflicts of interest in this work.

## References

Aeschlimann P, Häberli M, Reusch T, Boehm T, Milinski M (2003) Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology*, **54**, 119–126.

© 2016 John Wiley & Sons Ltd

Aloe AM (2015) Inaccuracy of regression results in replacing bivariate correlations. *Research Synthesis Methods*, **6**, 21–27.

Arnold C, Matthews LJ, Nunn CL (2010) The 10ktrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews*, **19**, 114–118.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bishop DVM, Thompson PA (2016) Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value. *PeerJ*, **4**, e1715.

Bovet J, Barthes J, Durand V, Raymond M, Alvergne A (2012) Men's preference for women's facial features: testing homogamy and the paternity uncertainty hypothesis. *PLoS ONE*, **7**, e49791.

Boyse EA, Beauchamp GK, Yamazaki K (1983) Critical review: the sensory perception of genotypic polymorphism of the major histocompatibility complex and other genes: some physiological and phylogenetic implications. *Human Immunology*, **6**, 177–183.

Brown JL (1997) A theory of mate choice based on heterozygosity. *Behavioral Ecology*, **8**, 60–65.

Brown JL (1999) The new heterozygosity theory of mate choice and the MHC. *Genetica*, **104**, 215–221.

Bruns SB, Ioannidis JPA (2016) *P-curve* and *p-hacking* in observational research. *PLoS ONE*, **11**, e0149144.

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *The Lancet*, **361**, 598–604.

Carroll LS, Penn DJ, Potts WK (2002) Discrimination of mhc-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 2187–2192.

Chaix RL, Cao C, Donnelly P (2008) Is mate choice in humans MHC-dependent? *PLoS Genetics*, **4**, e1000184.

Clutton-Brock TH (1989) Review lecture: mammalian mating systems. *Proceedings of the Royal Society of London: B: Biological Sciences*, **236**, 339–372.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. L. Erlbaum, Hillsdale, New Jersey.

Derti A, Cenik C, Kraft P, Roth FP (2010) Absence of evidence for MHC–dependent mate selection within hapmap populations. *PLoS Genetics*, **6**, e1000925.

Dixson AF, Anderson MJ (2004) Sexual behavior, reproductive physiology and sperm competition in male mammals. *Physiology & Behavior*, **83**, 361–371.

Duval S, Tweedie R (2000a) A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, **95**, 89–98.

Duval S, Tweedie R (2000b) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **56**, 455–463.

Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ*, **315**, 629–634.

Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, **18**, 3316–3329.

Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications*, **3**, doi: 10.1038/ncomms1632.

Ejsmond MJ, Radwan J, Wilson AB (2014) Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B: Biological Sciences*, **281**, doi: 10.1098/rspb.2014.1662.

Fernandez-Duque E, Valeggia CR, Mendoza SP (2009) The biology of paternal care in human and nonhuman primates. *Annual Review of Anthropology*, **38**, 115–130.

Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence handicap. *The American Naturalist*, **139**, 603–622.

Forsberg LA, Dannewitz J, Petersson E, Grahn M (2007) Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout – females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology*, **20**, 1859–1869.

Fromhage L, Kokko H, Reid JM (2009) Evolution of mate choice for genome-wide heterozygosity. *Evolution*, **63**, 684–694.

Fumagalli M, Sironi M, Pozzoli U *et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, **7**, doi: 10.1371/journal.pgen.1002355.

Gangestad SW, Buss DM (1993) Pathogen prevalence and human mate preferences. *Ethology and Sociobiology*, **14**, 89–96.

Garamszegi LZ (2014) Global distribution of malaria-resistant MHC-HLA alleles: the number and frequencies of alleles and malaria risk. *Malaria Journal*, **13**, doi: 10.1186/1475-2875-13-349.

Garamszegi LZ, Nunn CL (2011) Parasite-mediated evolution of the functional part of the MHC in primates. *Journal of Evolutionary Biology*, **24**, 184–195.

Garamszegi LZ, Calhim S, Dochtermann N *et al.* (2009) Changing philosophies and tools for statistical inferences in behavioral ecology. *Behavioral Ecology*, **20**, 1363–1375.

Giphart M, D'Amaro J (1983) HLA and reproduction? *Journal of immunogenetics*, **10**, 25–29.

Goulder PJR, Brander C, Tang Y *et al.* (2001) Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature*, **412**, 334–338.

Goüy de Bellocq J, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *Journal of Evolutionary Biology*, **21**, 1144–1150.

Gragert L, Madbouly A, Freeman J, Maiers M (2013) Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire us donor registry. *Human Immunology*, **74**, 1313–1320.

Gray AW, Boothroyd LG (2012) Female facial appearance and health. *Evolutionary Psychology*, **10**, doi: 10.1177/147470491201000108.

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

Hamilton WD, Zuk M (1982) Heritable true fitness and bright birds: a role for parasites? *Science*, **218**, 384–387.

Harcourt AH, Harvey PH, Larson SG, Short RV (1981) Testis weight, body weight and breeding system in primates. *Nature*, **293**, 55–57.

© 2016 John Wiley & Sons Ltd

Harcourt A, Purvis A, Liles L (1995) Sperm competition: mating system, not breeding season, affects testes size of primates. *Functional Ecology*, **9**, 468–476.

Havlicek J, Roberts SC (2009) MHC-correlated mate choice in humans: a review. *Psychoneuroendocrinology*, **34**, 497–512.

Hedges LV, Pigott TD (2001) The power of statistical tests in meta-analysis. *Psychological Methods*, **6**, 203–217.

Hedrick PW (1992) Female choice and variation in the major histocompatibility complex. *Genetics*, **132**, 575–581.

Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. *The American Naturalist*, **143**, 945–964.

Hedrick PW, Black FL (1997) HLA and mate selection: no evidence in South Amerindians. *American Journal of Human Genetics*, **61**, 505–511.

Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, **21**, 1539–1558.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557.

Hill AVS, Allsopp CEM, Kwiatkowski D *et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.

Hume DK, Montgomerie R (2001) Facial attractiveness signals different aspects of "quality" in women and men. *Evolution and Human Behavior*, **22**, 93–112.

Jacob S, McClintock MK, Zelano B, Ober C (2002) Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature Genetics*, **30**, 175–179.

Jiang Y, Bolnick DI, Kirkpatrick M (2013) Assortative mating in animals. *The American Naturalist*, **181**, E125–E138.

Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. *Heredity*, **81**, 127–133.

Kaessmann H, Wiebe V, Weiss G, Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genetics*, **27**, 155–156.

Kalmijn M (1998) Intermarriage and homogamy: causes, patterns, trends. *Annual Review of Sociology*, **000**, 395–421.

Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S (2014) A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Molecular Ecology*, **23**, 5151–5163.

Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, **15**, 379–393.

Kempenaers B (2007) Mate choice and genetic quality: a review of the heterozygosity theory. *Advances in the Study of Behavior*, **37**, 189–278.

Khankhanian P, Gourraud PA, Caillier SJ *et al.* (2010) Genetic variation in the odorant receptors family 13 and the MHC loci influence mate selection in a multiple sclerosis dataset. *BMC Genomics*, **11**, doi: 10.1186/1471-2164-11-626.

Knapp LA (2005) The ABCs of MHC. *Evolutionary Anthropology*, **14**, 28–37.

Lajeunesse MJ (2010) Achieving synthesis with meta-analysis by combining and comparing all available studies. *Ecology*, **91**, 2561–2564.

Langlois JH, Kalakanis L, Rubenstein AJ *et al.* (2000) Maxims or myths of beauty? a meta-analytic and theoretical review. *Psychological Bulletin*, **126**, 390–423.

Lashley LE, Haasnoot GW, Spruyt-Gerritse M, Claas FH (2015) Selective advantage of HLA matching in successful uncomplicated oocyte donation pregnancies. *Journal of Reproductive Immunology*, **112**, 29–33.

Laurent R, Chaix R (2012) MHC-dependent mate choice in humans: why genomic patterns from the hapmap European American dataset support the hypothesis. *BioEssays*, **34**, 267–271.

Laurent R, Toupance B, Chaix R (2012) Non-random mate choice in humans: insights from a genome scan. *Molecular Ecology*, **21**, 587–596.

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genetics*, **8**, e1002453.

Leinders-Zufall T, Brennan P, Widmayer P *et al.* (2004) MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science*, **306**, 1033–1037.

Lewis MB (2010) Why are mixed-race people perceived as more attractive? *Perception*, **39**, 136–138.

Lie HC, Rhodes G, Simmons LW (2008) Genetic diversity revealed in human faces. *Evolution*, **62**, 2473–2486.

Little AC, Hockings KJ, Apicella CL, Sousa C (2012) Mixed-ethnicity face shape and attractiveness in humans. *Perception*, **41**, 1486–1496.

Liu L, Johnson HL, Cousens S *et al.* (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, **379**, 2151–2161.

Magnusson A, Stewart I (2014) plotMCMC: MCMC Diagnostic Plots. R package version 2.0-0. https://CRAN.R-project.org/package=plotMCMC.

McClintock MK, Schumm P, Jacob S, Zelano B, Ober C (2002) Reply to "The MHC and body odors: arbitrary effects caused by shifts of mean pleasantness". *Nature Genetics*, **31**, 237–238.

McMichael A (2001) Human culture, ecological change, and infectious disease. *Ecosystem Health*, **7**, 107–115.

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics*, **37**, 159–186.

Milinski M, Bakker TCM (1990) Female sticklebacks use male coloration in mate choice and hence avoid parasitized males. *Nature*, **344**, 330–333.

Milinski M, Wedekind C (2001) Evidence for MHC-correlated perfume preferences in humans. *Behavioral Ecology*, **12**, 140–149.

Milinski M, Griffiths S, Wegner KM *et al.* (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4414–4418.

Milinski M, Croy I, Hummel T, Boehm T (2013) Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proceedings of the Royal Society B: Biological Sciences*, **280**, doi: 10.1098/rspb.2012.2889.

© 2016 John Wiley & Sons Ltd

Møller A, Jennions MD (2002) How much variance can be explained by ecologists and evolutionary biologists? *Oecologia*, **132**, 492–500.

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.

Nakagawa S, Santos ES (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, **26**, 1253–1274.

Niimura Y (2009) On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biology and Evolution*, **1**, 34–44.

Nojo S, Tamura S, Ihara Y (2012) Human homogamy in facial characteristics. *Human Nature*, **23**, 323–340.

Nowak MA, Tarczy-Hornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 10896–10899.

Ober C, Elias S, O'Brien E *et al.* (1988) HLA sharing and fertility in hutterite couples: evidence for prenatal selection against compatible fetuses. *American Journal of Reproductive Immunology and Microbiology: AJRIM*, **18**, 111–115.

Ober C, Weitkamp LR, Cox N *et al.* (1997) HLA and mate choice in humans. *American Journal of Human Genetics*, **61**, 497–504.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Paradis E, Claude J, Strimmer K (2004) Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, **20**, 289–290.

Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.

Penn D, Potts WK (1998) Chemical signals and parasite-mediated sexual selection. *Trends in Ecology & Evolution*, **13**, 391–396.

Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist*, **153**, 145–164.

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11260–11264.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Potts WK, Wakeland EK (1990) Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, **5**, 181–187.

Potts WK, Manning CJ, Wakeland EK (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature*, **352**, 619–621.

Potts WK, Manning CJ, Wakeland EK, Hughes AL (1994) The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **346**, 369–378.

Prado-Martinez J, Sudmant PH, Kidd JM *et al.* (2013) Great ape genetic diversity and population history. *Nature*, **499**, 471–475.

Prugnolle F, Manica A, Charpentier M *et al.* (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology*, **15**, 1022–1027.

Puts DA (2010) Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behavior*, **31**, 157–175.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rantala MJ, Moore FR, Skrinda I *et al.* (2012) Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature Communications*, **3**, 694. doi: 10.1038/ncomms1696.

Redden DT, Allison DB (2006) The effect of assortative mating upon genetic association studies: Spurious associations and population substructure in the absence of admixture. *Behavior Genetics*, **36**, 678–686.

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rhodes G, Lee K, Palermo R *et al.* (2005) Attractiveness of own-race, other-race, and mixed-race faces. *Perception*, **34**, 319–340.

Roberts SC (2009) Complexity and context of MHC-correlated mating preferences in wild populations. *Molecular Ecology*, **18**, 3121–3123.

Roberts SC, Little AC, Gosling LM *et al.* (2005a) MHC-assortative facial preferences in humans. *Biology Letters*, **1**, 400–403.

Roberts SC, Little AC, Gosling LM *et al.* (2005b) MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior*, **26**, 213–226.

Roberts SC, Little AC, Gosling LM, *et al.* (2005) MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior*, **26**, 213–226.

Roberts SC, Hale ML, Petrie M (2006) Correlations between heterozygosity and measures of genetic similarity: implications for understanding mate choice. *Journal of Evolutionary Biology*, **19**, 558–569.

Robinson J, Halliwell JA, Hayhurst JD *et al.* (2014) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, **43**, D423–D431.

Rodgers JR, Cook RG (2005) MHC class Ib molecules bridge innate and acquired immunity. *Nature Reviews Immunology*, **5**, 459–471.

Rosenberg LT, Cooperman D, Payn R (1983) HLA and mate selection. *Immunogenetics*, **17**, 89–93.

Rosenberg NA, Pritchard JK, Weber JL *et al.* (2002) Genetic structure of human populations. *Science*, **298**, 2381–2385.

Santos PSC, Schinemann JA, Gabardo J, da Graça BM (2005) New evidence that the MHC influences odor perception in humans: a study with 58 Southern Brazilian students. *Hormones and Behavior*, **47**, 384–388.

Setchell JM, Huchard E (2010) The hidden benefits of sex: evidence for MHC-associated mate choice in primate societies. *BioEssays*, **32**, 940–948.

Šimková A, Ottová E, Morand S (2006) MHC variability, life-traits and parasite diversity of European cyprinid fish. *Evolutionary Ecology*, **20**, 465–477.

Simonsohn U, Nelson LD, Simmons JP (2014a) P-curve and effect size correcting for publication bias using only significant results. *Perspectives on Psychological Science*, **9**, 666–681.

Simonsohn U, Nelson LD, Simmons JP (2014b) P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, **143**, 534–547.

Slade R, McCallum H (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, **132**, 861–862.

© 2016 John Wiley & Sons Ltd

Soetaert K (2014) shape: functions for plotting graphical shapes, colors. R package version 1.4.2. https://CRAN.R-project.org/package=shape.

Solberg OD, Mack SJ, Lancaster AK *et al.* (2008) Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Human Immunology*, **69**, 443–464.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639.

Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. *Trends in Ecology & Evolution*, **11**, 338–342.

Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology*, **20**, 4408–4420.

Taylor LH, Latham SM, Mark E (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**, 983–989.

Thomas L (1997) Retrospective power analysis. *Conservation Biology*, **11**, 276–280.

Thornhill R, Gangestad SW (1999) The scent of symmetry: a human sex pheromone that signals fitness? *Evolution and Human Behavior*, **20**, 175–201.

Thornhill R, Gangestad SW, Miller R *et al.* (2003) Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology*, **14**, 668–678.

Thoß M, Ilmonen P, Musolf K, Penn DJ (2011) Major histocompatibility complex heterozygosity enhances reproductive success. *Molecular Ecology*, **20**, 1546–1557.

Thursz MR, Kwiatkowski D, Allsopp CE *et al.* (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *New England Journal of Medicine*, **332**, 1065–1069.

Trachtenberg E, Korber B, Sollars C *et al.* (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, **9**, 928–935.

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular Ecology*, **9**, 1013–1027.

Trivers R (1972) Parental investment and sexual selection. In: *Sexual selection and the descent of man* (ed. Campbell B.), pp. 136–179. Aldine Publishing Company, Chicago, Illinois.

Tummers B (2006) Datathief iii. http://datathief.org.

Viechtbauer W (2010) Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, **36**, 1–48.

deVillemereuil P, Nakagawa S (2014) General quantitative genetic methods for comparative biology. In: *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pp. 287–303. Springer.

Vina MAF, Hollenbach JA, Lyke KE *et al.* (2012) Tracking human migrations by the analysis of the distribution of hla alleles, lineages and haplotypes in closed and open populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 820–829.

Wedekind C, Seebeck T, Bettens F, Paepke AJ (1995) MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **260**, 245–249.

Wedekind C, McClintock MK, Schumm P *et al.* (2002) The MHC and body odors: arbitrary effects caused by shifts of mean pleasantness. Authors' reply. *Nature Genetics*, **31**, 237–238.

Wegner K, Reusch T, Kalbe M (2003a) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16**, 224–232.

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003b) Parasite selection for immunogenetic optimality. *Science*, **301**, 1343.

Wilson AJ, Reale D, Clements MN *et al.* (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**, 13–26.

Winternitz J, Abbate J (2015) Examining the evidence for major histocompatibility complex-dependent mate selection in humans and nonhuman primates. *Research and Reports in Biology*, **2015**, 73–88.

Winternitz JC, Minchey SG, Garamszegi LZ *et al.* (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society B: Biological Sciences*, **280**, doi: 10.1098/rspb.2013.1605.

Woelfing B, Traulsen A, Milinski M, Boehm T (2009) Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 117–128.

Yamazaki K, Beauchamp GK, Kupniewski D *et al.* (1988) Familial imprinting determines H-2 selective mating preferences. *Science*, **240**, 1331–1332.

Yang B, Ren B, Xiang Z *et al.* (2014) Major histocompatibility complex and mate choice in the polygynous primate: the Sichuan snub-nosed monkey (*Rhinopithecus roxellana*). *Integrative Zoology*, **9**, 598–612.

J.W. initiated the study; J.W., J.A. and L.Z.G. designed the study; J.W., J.A. and L.Z.G. compiled data sets; J.W. and L.Z.G. analysed data; and J.W. wrote the manuscript with the substantial contribution of J.A., E.H., J.H. and L.Z.G.

## Data accessibility

The data sets compiled for our analyses and R code are deposited on Dryad (doi: 10.5061/dryad.5003g).

## Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Moderator estimates for MHC-dissimilar mating patterns.

**Table S2** Moderator estimates for human MHC-dissimilar mating patterns.

**Table S3** Moderator estimates for human MHC-dissimilar mating patterns by experimental design.

© 2016 John Wiley & Sons Ltd

**Table S4** Moderator estimates for non-human primate MHC-dissimilar mating patterns.

**Table S5** Population heterogeneity moderator estimates for human MHC-similar mating patterns.

**Table S6** Moderator estimates for MHC-diversity mating patterns.

**Table S7** Moderator estimates for human MHC-diversity mating patterns.

**Table S8** Moderator estimates for non-human primate MHC-diversity mating patterns.

**Table S9** Heterogeneity estimates for MHC-dissimilarity meta-regression models.

**Table S10** Heterogeneity estimates for MHC-diversity meta-regression models.

**Table S11** Moderator contrasts for MHC-dissimilarity.

**Table S12** Moderator contrasts for MHC-diversity.

**Table S13** Mixed-model results using alternative random effects.

**Table S14** Publication bias and sensitivity analysis results.

**Table S15** Testing for temporal bias in MHC mate choice datasets.

**Table S16** P-curve analysis results.

**Table S17** Power analysis.

**Fig. S1** Primate phylogenetic tree and mating system.

**Fig. S2** Publication bias and sensitivity analysis for MHC dissimilarity.

**Fig. S3** Publication bias and sensitivity analysis for MHC diversity.

**Fig. S4** Raw data for moderators of MHC-dissimilarity effect sizes (yi).

**Fig. S5** Raw data for moderators of human MHC-dissimilarity effect sizes (yi).

**Fig. S6** Raw data for moderators of non-human primate MHC-dissimilarity effect sizes (yi).

**Fig. S7** Raw data for moderators of MHC-diversity effect sizes (yi).

**Fig. S8** Raw data for moderators of human MHC-diversity effect sizes (yi).

**Fig. S9** Raw data for moderators of non-human primate MHC-diversity effect sizes (yi).

**Fig. S10** P-curve plots.

**Fig. S11** Correlation between MHC and neutral marker effect size for mate choice.

© 2016 John Wiley & Sons Ltd