Supplementary Material

SUPPLEMENTARY METHODS

Study species and Mapping populations

The genotyped individuals in the NL captive population were selected from 480 individuals of the fourth generation of artificial selection on Early Exploratory Behaviour (Drent *et al.* 2003) and their F1 and F2 intercrosses (van Oers *et al.* 2004). We bred F1 intercrosses from 2000 until 2008, and F2 intercrosses from 2002-2009. For the NL mapping population individuals were selected based on the pedigree links within the captive population. Offspring were checked for genetic compatibility with their genotyped parents by counting the number of Mendelian inheritance errors for autosomal SNPs using Genome Studio software (Illumina). When parent-offspring mismatches exceeded 100, offspring were omitted from the pedigree (N=7). For the final mapping pedigree (N=398), we included all F2 intercross individuals (N=251), their F1 parents, the siblings of these F1 parents and all full-sib selection line families with their parents. These individuals were manually split into 35 sub-families: 9 sub-families consisted of selection line offspring with their parents, 9 sub-families with F1 intercross offspring with their parents and 17 sub-families with F2 intercross offspring, their F1 intercross parents and their selection line grandparents. Three F2 families also contained half-sibs, and therefore included more than two F1 parents and more than four selection line grandparents. The mean family size was 15.0 (range 5-35). Sex was confirmed using the molecular method of Griffiths *et al*. (1998). We checked for non-inheritance errors using the PREPARE option in LISPCRI, a modified version of CriMap and excluded all non-inheritance errors by setting the genotypes to zero (thereby excluding them from further analyses) using a Perl script.

UK individuals were selected for genotyping based on (i) availability of sufficient DNA for genotyping (>50 ng/mL, as quantified by the PicoGreen assay), (ii) availability of phenotypic measurements, particularly morphological measurements, adult life history data, exploration rate and other personality data, social network data and basal metabolic rate measurements, and (iii) pedigree links within the Wytham pedigree,

which contains ~80,000 individuals born since 1958. The social pedigree of the genotyped individuals of the UK population was constructed from field observations, where expected parentage was determined from capturing males and females within nest boxes containing nestling great tits. Great tit nestlings (and previously un-marked adults) were marked individually using aluminium rings fitted to the tarsus. Relationships in the pedigree were confirmed using a number of methods. First, SNPs assigned a putative zebra finch mapping position on the Z chromosome (based on sequence homology between SNP flanking sequence and the zebra finch genome sequence) were used to confirm the sex of the genotyped individuals; homozygosity by locus (HL) of Z-linked markers was calculated using the 'GENHET' package (Coulon 2010) in R version 2.11.1 (R Development Core Team 2006). Individuals with HL values of 0-0.7 were assigned as male, while individuals with HL values of 0.9-1 were assigned as female, where the distribution of HL values is bimodal, with no overlap between the sexes. The ID of seven individuals whose genetic sex was incompatible with their recorded sex (i.e. sample mislabelling or lab errors) was assigned as "unknown" and all pedigree links to these samples were removed. Ten individuals with HL values between 0.7-0.9 were assigned their recorded sex (0.26% of individuals).

Second, genotyped individuals were checked for genetic compatibility with their genotyped parents and their offspring by counting the number of Mendelian inheritance errors for autosomal SNPs. Inheritance errors were summed for each parent individually to determine whether mismatches were predominantly due to mismatches with the social father, which is likely to be caused by extra pair paternity (i.e. inconsistency between the social and genetic pedigree), in which case the pedigree link to the father was removed (91 of the 682 individuals with both parents genotyped). In general, individuals who raised approximately equal numbers of mismatches with each parent were labelled as "unknown" and all pedigree links removed (40 of the 682 individuals with both parents genotyped), although in some cases (e.g. genotype consistency with the individual's offspring) pedigree links were only removed in one direction (i.e. links to the parents but not offspring). Where only one parent was genotyped and a large number of inheritance errors were raised, the individual was renamed as unknown and all pedigree links removed (81 of the 1,096 individuals with at least one parent genotyped). Next, identity by state (IBS) allele sharing at the autosomal markers was calculated between all pairs of individuals using PLINK (Purcell *et al.* 2007). This identified 72 pairs of individuals with identical genotypes (IBS values of >0.99), suggesting sample mislabelling or lab errors (representing 2.7% of the

genotyped individuals). In the absence of first-degree genotyped relatives to confirm the true identity of both individuals, both were labelled as "unknown" and pedigree links removed.

Finally, categorical paternity analyses were carried out to assign genetic fathers to individuals resulting from extra pair paternities (a total of 91 individuals compatible with their genotyped social mother but not their genotyped social father), using CERVUS 3.0 (Kalinowski *et al.* 2007). All genotyped males were included in the pool of candidate fathers. To assess whether the paternity could be assigned to the male with the highest paternity likelihood, the difference in likelihood score between the most likely fathers was compared to a simulated distribution obtained from 10,000 simulated mating events. Pedigree links were subsequently restored for the 59 birds (out of 91) for which a father was assigned with strict confidence (> 95%) in CERVUS. A subset of 1,656 Wytham individuals with confirmed first degree pedigree links (genotyped parents and/or sibs and/or offspring) were chosen for constructing the UK great tit linkage map. Parent-offspring genotype errors were removed by setting the genotypes to unknown. For the UK population complex pedigrees were first split into 61 subfamilies (of approximately 50 individuals and 3 generations) using the CRIGEN command (options: -size 50 and -gen 3) in the version of CriMap modified by Xuelu Liu (see below).

Map construction

For linkage mapping we used version 2.503 of the CriMap software (Green *et al.* 1990) modified by Jill Maddox (Department of Veterinary Science, University of Melbourne, Australia) to better handle large datasets , and its interface, the Linkage Mapping Software, developed by Xuelu Liu (Monsanto) to accommodate large numbers of markers segregating in a complicated pedigree structure. The likelihood of linkage between all possible pairs of markers was calculated using the TWOPOINT option (options phase unknown likelihood tolerance: PUK_LIKE_TOL = 1 and phase known likelihood tolerance: PK_LIKE_TOL = 1). Linkage mapping is an iterative process, and we chose to present 3 different steps of this process in this paper: (i) framework maps, incorporating markers when they could be assigned a most likely position with LOD > 3, where the best marker order is at least 10^3 = 1000 times more likely than any alternative order. (ii) comprehensive maps, where “non-framework” markers were added using lower levels of stringency (LOD > 0.1) and (iii) parsimonious maps, including as many additional markers as possible.

We applied two different initial strategies for linkage group construction. For the NL map, we assigned 6877 SNPs to putative great tit chromosomes based on sequence homology between SNP flanking sequences and the zebra finch genome sequence v 3.2.4.63. We used the option HAPSYS to group SNPs that had zero recombination with each other. Only the most informative SNP of such a haplotype-group was used for mapping. The recombination rates of the other SNPs were not fixed to zero, so that SNPs were not forced onto the same map position in cases where the recombination with other markers varied between markers within a haplotype group. This was done to lower calculation times, but will not result in a different map.

For the UK map, we used the TWOPOINT output of CriMap to create linkage groups with the AUTOGROUP command. AUTOGROUP employs an iterative process to group linked markers on the basis of the likelihood of linkage and the marker variability, beginning with very high stringency and proceeding through layers of lower stringency. Parameters include (i) the minimum threshold of LOD score for a linkage to be considered, (ii) minimum number of informative meioses for a marker to be included in terms of x times the mean number of informative meioses, (iii) the maximum number of linkages to other groups, and (iv) minimum linkage ratio for a marker's qualified links to the best linkage group (i.e., the proportion of two-point linkages for a given marker that are to markers that are all in the same linkage group) (see e.g. Stapley *et al.* 2008). The parameter layers were as follows: layer 1 (40, 2.0, 2.0, 0.9); layer 2 (20, 1.5, 2, 0.9); layer 3 (10, 1.0, 2, 0.9); and layer 4 (10, 0.5, 2, 0.9).

(i) Building framework maps

To construct both the UK and NL chromosome maps, the most parsimonious marker order was determined using the BUILD command. BUILD uses an iterative multipoint-likelihood maximization process, thereby producing the order with the highest likelihood. In general, we choose two or three markers as anchors for the first build, selecting markers on the basis of a high number of informative meiosis, intermediate recombination rates and high twopoint LOD scores. For the NL map, markers were initially added to the map with very high stringency (options PUK_LIKE_TOL = 7 and PK_LIKE_TOL = 7). The stringency was then lowered stepwise to PUK_LIKE_TOL = 3 and PK_LIKE_TOL = 3. For construction of the UK framework map, markers were added to the map with 'high' stringency (options PUK_LIKE_TOL = 3 and PK_LIKE_TOL = 3) in one step. These LOD-score

values (>3) indicate that markers were only added to a map position if that solution was 1,000 times more likely than any other position.

Next, the order of markers in the framework map was tested by running the FLIPS7 option, which carries out permutations of marker order within groups of seven consecutive markers and calculates the likelihood of all other possible orders. If the initial likelihood was improved by locally reordering the markers, this change was made and FLIPS7 run again until no alternative order with higher likelihood could be found. Additionally, markers causing high numbers of double recombinants (i.e., markers likely to have relatively high error rates and/or in an incorrect position in the map) were identified using CHROMPIC. The modified version of CriMap (2.5.03) produces an output table with possible erroneous genotypes. For the NL dataset, we removed potentially problematic markers and excluded incorrect genotypes with a Python script (www.python.org) we wrote to update the genotype files, while for the UK dataset a script was written to identify problematic markers from the CHROMPIC output.

After this initial step, the process of building the map was repeated, now starting with the last build order. BUILD, FLIPS7 (or FLIPS5, with increasing marker numbers for the NL dataset) and CHROMPIC steps were repeated until no bad markers and no better order appeared.

(ii) Building comprehensive maps

Once framework maps were completed, comprehensive maps were built for each linkage group. Framework maps of each population were used as anchors to start adding more markers to the map. Markers were added stepwise by lowering PUK_LIKE_TOL and PK_LIKE_TOL to 0.1; as with the framework maps, this was done in one step for the UK map and in multiple steps for the NL map. FLIPS5 and CHROMPIC options were then used to check for local rearrangements or bad markers or genotypes. Again, BUILD, FLIPS5 and CHROMPIC were all repeated until no better orders were identified.

(iii) Building parsimonious maps

A final stage of mapping was carried out to add as many markers as possible to the **comprehensive** map (parsimonious map). PUK_LIKE_TOL and PK_LIKE_TOL were lowered to 0.001 to identify markers to add to the

existing comprehensive maps. For the UK map, new markers (compared to the comprehensive map) were discarded if they aligned to a different zebra finch chromosome than the other markers on the linkage group (this was not applicable for the NL map, since they were already discarded in the first step). For both for the UK and NL map, markers were discarded if their predicted zebra finch physical position was very different (> 5 000 000 bases) from other markers next to the mapping position of the parsimonious marker had been added to. For the NL map, after lowering PUK_LIKE_TOL and PK_LIKE_TOL to 0.001, the only remaining unmapped markers had multiple possible positions on the map. From these, markers that could be positioned to fewer than four possible adjacent map positions with equal likelihood were added to the NL comprehensive map. For the UK map, the marker was added next to the marker with the closest position on the zebra finch genome, provided the map location of the latter marker was consistent with its location on the zebra finch genome. Markers that increased the overall length in centiMorgans (cM) of the chromosome substantially (~10% or more), were discarded. Map figures were drawn using the MapChart software version 2.2 (Voorrips 2002).

For the UK data set, the entire mapping process took over six months for all chromosomes to be completed, with FLIPS5 on the comprehensive maps in particular taking up to four weeks for the larger chromosomes. For the NL linkage maps, this process was longer (over 14 months) due to the stepwise lowering of the likelihood thresholds, limitations on computing memory, and one (KvO) versus two (AWS & IDC) personnel dedicated to map construction.

Coulon A. (2010). GENHET: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources*, 10, 167-169.

Drent P.J., van Oers K. & van Noordwijk A.J. (2003). Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270, 45-51.

Green P., Falls K. & Crooks S. (1990). Documentation for CRI-MAP. In. Washington University, St. Louis.

Griffiths R., Double M.C., Orr K. & Dawson J.G. (1998). A DNA test to sex most birds. *Molecular Ecology*, 7, 1071-1075.

Kalinowski S.T., Taper M.L. & Marshall T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16, 1099-1106.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J. & Sham P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559-575.

Stapley J., Birkhead T.R., Burke T. & Slate J. (2008). A linkage map of the zebra finch Taeniopygia guttata provides new insights into avian genome evolution. *Genetics*, 179, 651-667.

Team R.D.C. (2006). R Foundation for Statistical Computing. URL http://www.R-project.org. In. R foundation for Statistical Computing Vienna, Austria.

van Oers K., Drent P.J., De Jong G. & van Noordwijk A.J. (2004). Additive and nonadditive genetic variation in avian personality traits. *Heredity*, 93, 496-503.

Voorrips R.E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, 93, 77-78.

Titles and legends to Supplementary figures

Supplementary Table 1. The between-population cross-correlation in local heterochiasmy, when SDI_local was estimated in 20-marker intervals in a 5-marker sliding window analysis.

Supplementary figure 1. Parsimonious genetic linkage map in centiMorgan (cM) of the great tit mapped in the NL population. The framework map loci (order supported by LOD>3) are in boldface and underlined on the linkage maps. **Of these, framework markers that were unique to the NL map are indicated in black. SNPs that were present in the same order on the framework maps of both the NL and the UK population are indicated in blue, framework markers that that were present on both maps, but in different order are indicated in red**.

Supplementary figure 2. Parsimonious genetic linkage map in centiMorgan (cM) of the great tit mapped in the UK population. The framework map loci (order supported by LOD>3) are in boldface and underlined on the linkage maps. **Of these, framework markers that were unique to the UK map are indicated in black. SNPs that were present in the same order on the framework maps of both the NL and the UK population are indicated in blue, framework markers that that were present on both maps, but in different order are indicated in red**.

Supplementary figure 3. The parsimonious linkage map positions in centiMorgan (cM) of markers that were mapped on both the UK and the NL linkage map.

Supplementary figure 4. Recombination landscape for each great tit linkage map on the NL linkage map, plotted as the linkage map position in centiMorgan (cM) against the map order.

Supplementary figure 5. **Size Dimorphism Index (SDI) calculated for windows of 20 (a), 30 (b) or 50 (c) SNP markers on the NL (orange) and UK (blue) framework maps. Subsequent windows were chosen by sliding the window 10 (a), 15 (b) or 10 (c) SNP markers along the linkage group. Positive SDI estimates indicate that female recombination rates of that 20 SNP marker window was higher, negative values when the male rates are highest. Only linkage groups are plotted with at least three windows or more.**

Supplementary figure 6. Comparative maps containing great tit linkage map positions in centiMorgans (cM) plotted against the predicted zebra finch physical position for each great tit chromosome. Only chromosomes with predicted zebra finch positions for 10 or more markers are plotted.

Supplementary figure 7. Comparative maps containing great tit linkage map positions in centiMorgans (cM) plotted against the predicted chicken physical position for each great tit chromosome. Only chromosomes with predicted chicken positions for 10 or more markers are plotted.

Supplementary table 1.

Linkage group	Cross Correlation in SDI_local
1	0.668
1A	0.870
2	0.375
3	0.477
4	0.646
4A	0.494
5	0.813
6	0.050
7	0.649
8	0.459
9	0.598
10	0.575
11	0.621
12	0.815
13	0.712
14	0.690
15	0.553
17	0.673
18	0.919
19	0.963
20	0.694
21	0.999
24	0.992