View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Kent Academic Repository

Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution

Martin Völker, Niclas Backström, Benjamin M. Skinner, et al.

Genome Res. 2010 20: 503-511 originally published online March 31, 2010
Access the most recent version at doi:10.1101/gr.103663.109

Supplemental Material http://genome.cshlp.org/content/suppl/2010/02/09/gr.103663.109.DC1.html

Related Content

Sex bias and dosage compensation in the zebra finch versus chicken genomes: General and specialized patterns among birds
Yuichiro Itoh, Kirstin Replogle, Yong-Hwan Kim, et al.
Genome Res. April , 2010 20: 512-518 **The recombination landscape of the zebra finch Taeniopygia guttata genome**
Niclas Backström, Wolfgang Forstmeier, Holger Schielzeth, et al.
Genome Res. April , 2010 20: 485-495 **Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome**
Jessica Stapley, Tim R. Birkhead, Terry Burke, et al.
Genome Res. April , 2010 20: 496-502

References This article cites 62 articles, 16 of which can be accessed free at:
http://genome.cshlp.org/content/20/4/503.full.html#ref-list-1

Articles cited in:
http://genome.cshlp.org/content/20/4/503.full.html#related-urls

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

Copyright © 2010 by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution

Martin Völker,[1] Niclas Backström,[2] Benjamin M. Skinner,[1,3] Elizabeth J. Langley,[1,4] Sydney K. Bunzey,[1] Hans Ellegren,[2] and Darren K. Griffin[1,5]

[1]*School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, United Kingdom;* [2]*Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden*

Chromosomal rearrangements and copy number variants (CNVs) play key roles in genome evolution and genetic disease; however, the molecular mechanisms underlying these types of structural genomic variation are not fully understood. The availability of complete genome sequences for two bird species, the chicken and the zebra finch, provides, for the first time, an ideal opportunity to analyze the relationship between structural genomic variation (chromosomal and CNV) and recombination on a genome-wide level. The aims of this study were therefore threefold: (1) to combine bioinformatics, physical mapping to produce comprehensive comparative maps of the genomes of chicken and zebra finch. In so doing, this allowed the identification of evolutionary chromosomal rearrangements distinguishing them. The previously reported interchromosomal conservation of synteny was confirmed, but a larger than expected number of intrachromosomal rearrangements were reported; (2) to hybridize zebra finch genomic DNA to a chicken tiling path microarray and identify CNVs in the zebra finch genome relative to chicken; 32 interspecific CNVs were identified; and (3) to test the hypothesis that there is an association between CNV, chromosomal rearrangements, and recombination by correlating data from (1) and (2) with recombination rate data from a high-resolution genetic linkage map of the zebra finch. We found a highly significant association of both chromosomal rearrangements and CNVs with elevated recombination rates. The results thus provide support for the notion of recombination-based processes playing a major role in avian genome evolution.

[Supplemental material is available online at http://www.genome.org. The array CGH data from this study have been submitted to ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) under accession no. E-TABM-841.]

Large-scale variations in genomic structure include chromosomal rearrangements and copy number variants (CNVs)—lengths of DNA 1 kb or larger, present in variable copy number compared to a reference genome. Both have been strongly implicated as playing an important role in phenotypic variation, genetic disease, and genome evolution (e.g., Shaw and Lupski 2004; Coghlan et al. 2005; Bailey and Eichler 2006; Sharp et al. 2006; Ferguson-Smith and Trifonov 2007; Hoffmann and Rieseberg 2008; Hurles et al. 2008; Manolio et al. 2009; Zhang et al. 2009).

The full extent and impact of CNVs (in the human and other genomes) is becoming increasingly apparent due to the availability of complete genome sequence assemblies, high-throughput detection platforms (i.e., microarrays), and robust bioinformatics analysis protocols. For instance, there is now evidence that CNVs encompass more nucleotide content per human genome than single nucleotide polymorphisms (SNPs) (Redon et al. 2006); that CNVs are associated with numerous human diseases (Wain et al. 2009; Zhang et al. 2009); and that CNVs play a role in evolutionary adaptation (Perry et al. 2007; Nair et al. 2008). Moreover, studies in primates have revealed many lineage-specific CNVs, including gene copy number changes (Cheng et al. 2005; Perry et al. 2006; Dumas et al. 2007; Marques-Bonet et al. 2009).

Present addresses: [3]Mammalian Molecular Genetics Group, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK; [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus Hinxton, Cambridge CB10 1SA, UK.
[5]Corresponding author.
E-mail D.K.Griffin@kent.ac.uk; fax 44 (0) 1227-763912.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.103663.109.

The medical and evolutionary importance of large-scale structural chromosomal rearrangements is much more well-established; they underlie numerous human diseases, specifically cancer, birth defects, and infertility (Online Chromosomal Variation in Man, http://www.wiley.com/legacy/products/subject/life/borgaonkar/access.html). They also can cause or reinforce reproductive isolation between species (Noor et al. 2001; Rieseberg 2001; Delneri et al. 2003).

The underlying causes of structural genomic variation are not entirely understood; however, the available data suggest a role for recombination-based mechanisms. Specifically, non-allelic homologous recombination (NAHR; also known as "ectopic recombination") and break-induced replication (BIR) are thought to be major causes of both CNVs and chromosomal rearrangements (Stankiewicz and Lupski 2002; Lupski and Stankiewicz 2005; Gu et al. 2008; Hastings et al. 2009). Both NAHR and BIR are based on homologous recombination, and both can result in CNVs and/or chromosomal rearrangements if recombination occurs between homologous sequences in different (non-allelic) chromosomal positions. Further support for the suggested link between NAHR/BIR and structural genomic variation comes from studies in humans, other primates, mice, dogs, and *Plasmodium falciparum*. These studies reported a significant correlation between CNVs and segmental duplications, (i.e., large, low copy repeats up to 400 kb in size that are thought to provide the substrate for NAHR) (Locke et al. 2003; Sharp et al. 2005; Goidts et al. 2006; Mok et al. 2008; Perry et al. 2008; She et al. 2008; Nicholas et al. 2009). Moreover, some disease-associated chromosomal rearrangements in humans occur in recombination hotspots (Visser et al. 2005; Lindsay et al. 2006; Myers et al. 2008). Taken together, the evidence for a relationship between recombination and

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

structural genomic variation is strong; however, to the best of our knowledge, it has not been ascertained fully in an evolutionary context, i.e., between species. That is, if recombination were indeed a driver of structural genomic change, and if the recombination landscape has remained conserved during evolution, then we would expect to see an enrichment of CNVs and chromosomal rearrangements in genomic regions with elevated recombination rates. The purpose of this study is to test this hypothesis using two recently sequenced bird genomes as models.

Bird genomes possess a unique combination of features that make them particularly suitable for analysis of the relationship between CNVs, chromosomal rearrangements, and recombination. Firstly, they have a small and conserved size; that is, the mean size of bird genomes and the range of variation in genome size are the smallest observed in any vertebrate group (http://www.genomesize.com). Data from the chicken genome sequencing project show that on a molecular level, the small size is due to a lower number of interspersed repeats, segmental duplications, and pseudogenes compared to mammalian genomes, despite a similar number of genes (The International Chicken Genome Sequencing Consortium 2004). Similarly, overall karyotype structure is highly conserved in birds, with the vast majority of extant species displaying a karyotype with about $2n = 80$ chromosomes, comprising few macro- and many small microchromosomes, and a ZZ/ZW sex chromosome pair (Christidis 1990; Rodionov 1997; Griffin et al. 2007). In line with the above data, recent studies of CNVs in turkey and duck (relative to chicken) have suggested that bird genomes also contain a low number of CNVs compared to mammalian genomes (Griffin et al. 2008; Skinner et al. 2009). Recombination rates in chicken and zebra finch, the only two bird species for which there is information on absolute rates of recombination (The International Chicken Genome Sequencing Consortium 2004; Groenen et al. 2009; Backström et al. 2010), are higher overall than in mammals. They also show substantial variation within the same genome, particularly between macro- and microchromosomes. Moreover, the evolutionary conservation of avian genomes in terms of size, overall karyotype structure, and limited structural variation forms a distinct contrast with the situation in mammals (Ferguson-Smith and Trifonov 2007). Taking all the above points together, we would therefore contend that studies in birds are particularly suitable as a starting point to ask whether hypotheses pertaining to genome dynamics derived from studies in mammals apply generally across the animal kingdom.

The availability of complete genome sequences for two bird species, the chicken (*Gallus gallus*, GGA; The International Chicken Genome Sequencing Consortium 2004) and zebra finch (*Taeniopygia guttata*, TGU) (Warren et al. 2010), provides, for the first time, an ideal opportunity to analyze the relationship between structural genomic variation (chromosomal and CNV) and recombination on a genome-wide level between two avian species. The aims of this study were therefore:

1. To combine bioinformatics, synteny analyses, and physical mapping by fluorescent in situ hybridization (FISH) to produce comprehensive comparative maps of the genomes of chicken and zebra finch. In so doing, this allowed the identification of evolutionary chromosomal rearrangements distinguishing them. Previous chromosome painting studies demonstrated a high degree of conserved synteny in chicken and zebra finch macrochromosomes, with only two interchromosomal rearrangements (GGA1 has undergone a fission that produced TGU1 and TGU1A, and GGA4 is present as two separate chromosomes, TGU4 and TGU4A, in the zebra finch genome) (Itoh and Arnold 2005). Despite this, a genetic linkage mapping study has suggested numerous intrachromosomal rearrangements between the two species (Stapley et al. 2008).
2. To hybridize zebra finch genomic DNA to a chicken whole-genome tiling path array to identify and locate CNVs in the zebra finch genome relative to chicken.
3. To test the hypothesis that there is an association between CNV, chromosomal rearrangements, and recombination by correlating data from (1) and (2) with recombination rate data from a high-resolution genetic linkage map of the zebra finch (Backström et al. 2010).

Results

Comprehensive comparative genomic maps between chicken and zebra finch reveal chromosomal rearrangements and conserved synteny

With respect to aim (1) in the introduction, whole-chromosome alignments of draft genome sequences confirmed previous results demonstrating a high degree of conserved synteny in the macrochromosomes, with only two *inter*chromosomal rearrangements distinguishing the chicken and zebra finch genomes (Itoh and Arnold 2005). Our analysis suggested that synteny is also conserved in the microchromosomes, with the possible exception of chicken chromosome 16 and its zebra finch ortholog. This chromosome bears the major histocompatibility complex (MHC) genes in chicken and may have undergone a fission in the zebra finch (CN Balakrishnan, R Ekblom, M Völker, H Westerdahl, R Godinez, H Kotkiewicz, DW Burt, T Graves, DK Griffin, WC Warren, et al., unpubl.). Alignment of whole-chromosome sequences of orthologous chicken and zebra finch chromosomes to visualize large-scale *intra*chromosomal rearrangements, however, identified a large number of differences. In total, 114 tentative intrachromosomal rearrangements (56 inversions and 58 translocations) were detected in all macro- and microchromosomes except chicken chromosomes 17 and 19. Figure 1 illustrates these findings for chicken chromosome 4 and its zebra finch orthologs "4" and "4A." Due to the small amount of sequence available for chicken chromosomes 16 and W (and their zebra finch orthologs), these chromosomes could not be analyzed.

Refinement of the above analysis by comparing the order of 10,499 orthologous gene pairs in the chicken and zebra finch genomes identified a total of 199 homologous synteny blocks (HSBs) ranging in size from 23 kb to 77,507 kb with mean and median sizes of 5015 kb and 1782 kb, respectively (Supplemental Table 1). Three of these HSBs were nested within other HSBs and six overlapped with other HSBs. For subsequent analyses, overlapping HSBs were merged, and nested HSBs were treated as independent HSBs (see Methods for details), leaving a remainder of 193 HSBs. The genomic location of HSBs was consistent with the rearrangements suggested by the alignment of whole-chromosome sequences.

In order to validate these results, we selected a subset of chromosomal rearrangements on chicken chromosomes 1–8 and Z and their zebra finch orthologs for physical mapping by FISH with 131 chicken and 131 zebra finch bacterial artificial chromosomes (BACs) containing orthologous sequences (Supplemental Table 2). Figure 1 also summarizes the results of the sequence alignments and the FISH mapping experiments for chicken chromosome 4 and the orthologous chromosomes 4 and 4A, which revealed several

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

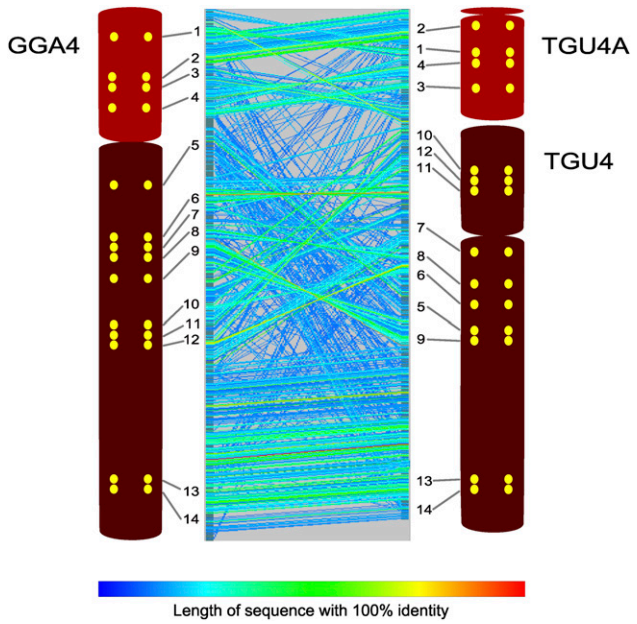

Figure 1. Comparative analysis of marker order on chicken chromosome 4 (GGA4) and its zebra finch orthologs, TGU4 and TGU4A. The central part of the figure was created by aligning whole-chromosome sequences using the program GenAlyzer (Choudhuri et al. 2004). Line color between chromosome bars indicates the lengths of sequences exhibiting 100% sequence identity (between 100 and 900 bp). The rearrangements suggested by this analysis were verified using fluorescent in situ hybridization (FISH). Numbers indicate the positions of chicken and zebra finch bacterial artificial chromosome (BAC) clones with orthologous sequence content in the genome sequences of both species. Yellow dots illustrate the physical position as determined by FISH.

inversions and translocations as well as a large region of conserved linkage on the distal part. Overall, the results of sequence alignments and FISH mapping were in close agreement, which provides independent cytogenetic evidence for the accuracy of the chicken and zebra finch genome sequence assemblies.

The small size of avian microchromosomes precluded the analysis of marker order by FISH on metaphase chromosomes. Therefore, the comparative FISH-based analysis of these chromosomes was restricted to the examination of conserved synteny by dual-color FISH with two BACs assigned to the same microchromosomes in the chicken and zebra finch genome assemblies. The FISH experiments confirmed the conserved synteny suggested by the bioinformatics analysis for all microchromosomes for which probes were available (chicken chromosomes 9–15, 16–28, and zebra finch chromosomes 9–15, 16–28). Figure 2 illustrates this finding for orthologous chicken/zebra finch chromosomes 24.

Relative CNVs in the zebra finch compared to chicken

With respect to aim (2) in the introduction, we detected relative CNVs in the zebra finch (compared to chicken) following the approach previously established for the analysis of CNVs in turkey (Griffin et al. 2008) and duck (Skinner et al. 2009), i.e., by performing array comparative genomic hybridization (CGH) with zebra finch genomic DNA samples on a chicken whole-genome tiling path microarray. In the Discussion and Methods sections, we have outlined the caveats pertaining to how sequence divergence between the two species may lead us to call an apparent loss in copy number by this approach; however, correct identification of Z chromosome copy number in male/female hybridizations provided independent validation of hybridization success. We detected 19 CNVs in the male zebra finch and 13 CNVs in the female zebra finch (Table 1). A total of 17 CNVs (63%) showed gains relative to chicken (i.e., suggesting that there were more copies of this sequence in zebra finch than chicken) and 10 CNVs (37%) contained losses (implying fewer copies relative to chicken). Overlapping CNVs found in both individuals analyzed (i.e., CNVs at the same locus in both birds) were, for the purposes of this study, defined as copy number variable regions (CNVRs). According to this definition, we detected five CNVRs: Four of these indicated relatively more copies of this locus in both zebra finches compared to the chicken reference; while in the fifth, the relative fluorescent ratios were consistent with either sequence loss, sequence divergence, or fewer copies in zebra finch compared to chicken.

In order to determine the genomic coordinates of CNVs in the zebra finch genome, the entire chicken sequence contained within a given CNV was aligned against the zebra finch genome using the BLAT algorithm. This analysis, which takes chromosomal rearrangements and sequence divergence into account, returned best hits on the expected orthologous zebra finch chromosome for 22 out of 27 CNVs. Of the remaining five CNVs, number 22 did not produce any hit; numbers 4, 8, and 27 produced hits on several chromosomes; and number 5 returned a hit on a different chromosome from the one expected (Table 1); we excluded these from further analyses. All five CNVs that failed to give unequivocal BLAT results contained losses in zebra finch relative to chicken.

The genomic coordinates of CNVs in the chicken and zebra finch genomes allowed for a comparison of CNV and CNVR sizes as estimated by array and alignment data. The sizes estimated from the array data ranged from 20.1 kb to 1225 kb, with mean and median sizes of 246.1 kb and 127.1 kb, respectively, while the sizes estimated from the BLAT alignments ranged from 12.6 kb to 1488.1 kb, with mean and median sizes of 314.1 kb and 142.9 kb, respectively. The sizes estimated from array and alignment data were highly correlated (Spearman's rank correlation, $\rho = 0.97$, $n = 22$, $P < 0.01$).

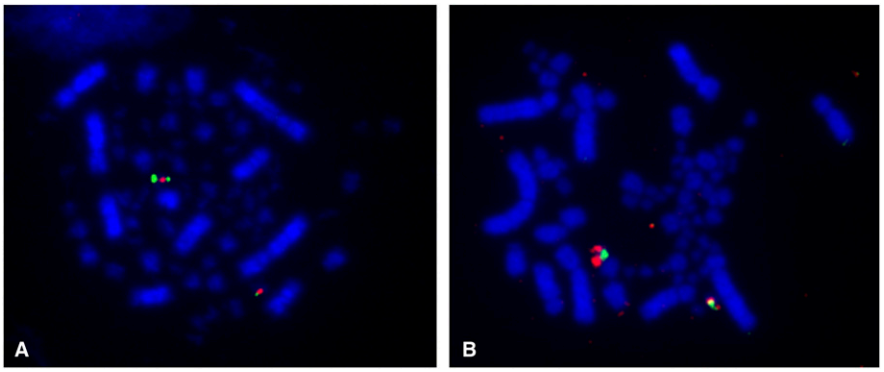

Figure 2. Analysis of conserved synteny in chicken (Griffin et al. 2008) and zebra finch (TGU) microchromosomes by dual color fluorescent in situ hybridization (FISH) using chicken and zebra finch bacterial artificial chromosome (BAC) clones with orthologous sequence content. (*A*) Chicken BACs WAG-20E08 (red) and WAG-21I18 (green) on GGA24. (*B*) The corresponding zebra finch BACs TGAC-82A15 (red) and TGAC-321M03 on TGU24.

Table 1. CNVRs and associated genes in the chicken (*Gallus gallus*, GGA) and zebra finch (*Taeniopygia guttata*, TGU) genomes

CNVR ID	GGA chr	GGA start	GGA end	GGA CNVR size (Balakrishnan 2010)	Genes associated in GGA	TGU chr	TGU start	TGU end	TGU CNVR size (Balakrishnan et al. 2010)	Status[c]	Male/ female	Genes associated in TGU
1	1	26847993	26917645	69.652	*CAV1_CHICK, NP_001007087.1*	1A	24499426	24575089	75.663	Loss	M	*CAV1*, *CAV2*
2	1	169725133	169752564	27.431	None	1	59991025	60017661	26.636	Loss	M	None
3	10	3912500	3937500	25.000	Novel	10	1281654	1309892	28.238	Gain	M	Novel
4[a]	11	2675046	3190227	515.181	None	Un, 6, 22	n/a	n/a	n/a	Loss	F	n/a
5	12	15208	42826	27.618	*LIMS1*	1	27277113	27293559	16.446	Loss	F	*XP_002192376.1*, Novel
6[a,b]	13	18780447	18907544	127.097	*EGR1_CHICK*	13	12013	83559	71.546	Loss	F	*EGR1_TAEGU*, *REEP2*, *JMJD1B*
7	15	917544	937693	20.149	*NP_989796.1*	15	7883666	7899849	16.183	Gain	M	*TBX1*
8[a,b]	16	12500	432851	420.351	Most of GGA16 sequence	Multiple	n/a	n/a	n/a	Loss	M+F	n/a
9	17	10137709	10170150	32.441	Novel	17	10623696	10656088	32.392	Gain	M	*GPR144*, *NR5A1*
10	18	9565043	9857515	292.472	*ENPP7*, *CBX8*, *NP_989973.1*, *TBC1D16*, *CBX2*, *C1QTNF1*, *Q8JGT3_CHICK*	18	1403912	1727088	323.176	Gain	M+F	Multiple, known and novel
11	2	177514	487500	309.986	*AGAP3*	2	146144	545934	399.790	Gain	M	Multiple, known and novel incl. *AGAP3*
12	20	9712500	10662500	950.000	Multiple, known and novel	20	7344763	8513526	1168.763	Gain	M	Multiple, known and novel
13	23	1077949	1307784	229.835	*SCMH1*, *CTPS*	23	1567394	2055849	488.455	Gain	M	*SCMH1*, *CTPS*
14	23	5500034	5887500	387.466	*SYYC_CHICK*, *MACF1*, *PPIE*, *ADC*	23	3427478	4063269	635.791	Gain	M+F	Multiple, known and novel
15	28	2900219	2945357	45.138	*CRTC1*	28	4866671	4911165	44.494	Loss	F	*CRTC1*
16	3	31462500	31637500	175.000	*CANX_CHICK*, *HS90B_CHICK*	3	31112716	31353318	240.602	Gain	M	Multiple, known and novel
17	5	12500	62500	50.000	*Q5ZMT2_CHICK*	5	7993275	8038577	45.302	Gain	M+F	*PLEKHG3*
18	5	300133	395182	95.049	*DDB1_CHICK*, *XR_026960.1*, *PEPA_CHICK*, *NP_990252.1*	5	6640346	6783297	142.951	Gain	M	Multiple, known and novel
19	6	23650363	23970419	320.056	*LCOR*, *SLIT1*, *RHG19_CHICK*, *MMS19*, *LOXL4*, *RRP12_CHICK*	6	21377425	21810043	432.618	Gain	M	Multiple, known and novel
20[b]	7	23630409	24037500	407.091	*INHA_CHICK*, *NP_001006590.1*, *NP_990260.1*, *TBA5_CHICK*, *NM_204179.1*, *NP_001012937.1*	7	10171925	10791843	619.918	Gain	M+F	Multiple, known and novel
21[a]	7	38335030	38380092	45.062	*NAT5*	7	39813464	39826002	12.538	Loss	F	*MARCH7*
22[a]	8	102	22657	22.555	None	No hit	n/a	n/a	n/a	Loss	F	n/a
23	8	21037500	21512500	475.000	*IPO13*, *PRNPIP*, *ATP6V0B*, *DMAP1*, *NM_001031279.1*	8	17150747	17752518	601.771	Gain	M	Multiple, known and novel
24	9	12720374	12770046	49.672	None	9	13163945	13221974	58.029	Gain	F	None
25	9	16962500	17137500	175.000	*NP_001012952.1*, *EPHB3_CHICK*, *Q5ZIM4_CHICK*	9	18143554	18398961	255.407	Gain	M	Multiple, known and novel
26	Z	7362500	8587500	1225.000	Multiple, known and novel	Z	38481856	39969939	1488.083	Gain	M	Multiple, known and novel
27[a,b]	Z	71687500	71812500	125.000	Novel	Multiple	n/a	n/a	n/a	Loss	F	n/a

[a]CNVRs also found in duck, *Anas platyrhynchos* (Skinner et al. 2009).
[b]CNVRs also found in turkey, *Meleagris gallopavo* (Griffin et al. 2008).
[c]The Status column indicates whether the zebra finch genome contained more copies (gain) or fewer copies (loss) of the DNA sequence contained within a given CNVR compared to the reference (chicken) genome.
n/a, Not available.

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

A total of 21 out of 22 CNVs (95%) with a plausible physical location in the zebra finch genome according to BLAT alignments were associated with genes, while 23 out of the total 27 CNVs (85%) were associated with genes in the chicken genome sequence. Six of the 27 CNVs had been previously detected in duck, four had been found in turkey, and three in both (Griffin et al. 2008; Skinner et al. 2009).

The genomic positions of both chromosomal breakpoints and CNVs correlate to regions of high genetic recombination

In order to determine if there was an association between recombination rate and chromosomal rearrangements (aim [3] in the introduction), we compared recombination rates in 118 1-Mb nonoverlapping windows with chromosomal breakpoints with recombination rates in 643 windows without breakpoints. The recombination rates were estimated from zebra finch pedigrees (Backström et al. 2010). This analysis revealed a significantly higher recombination rate in windows with chromosomal breakpoints (mean ± SD = 2.13 ± 3.03 cM/Mb) than in windows without breakpoints (1.14 ± 1.95 cM/Mb; Wilcoxon's rank sum test, W = 6376, P = 0.0000183; Fig. 3).

Likewise, we tested for an association between recombination rate and CNVs by comparing the recombination rate for 17 CNV-containing windows (median = 3.24 cM/Mb) with the recombination rate in the remaining 741 windows (1.24 cM/Mb) (Table 2). Again, the observed difference was statistically significant (Randomization test, P-value = 0.033; Fig. 4). Although not verified with statistical significance, a more detailed analysis indicated that windows containing both chromosomal breakpoints and CNVs tend to have a higher recombination rate (n = 5, median = 1.34 cM/Mb) than windows with breakpoints only (n = 111, 0.47 cM/Mb; Randomization test, P-value = 0.13) or CNVs only (n = 12, 0.69 cM/Mb, P-value = 0.081).

Discussion

To the best of our knowledge, the present study provides the first genome-wide analysis of correlation between chromosomal rearrangements, CNV, and recombination in any vertebrate genome.

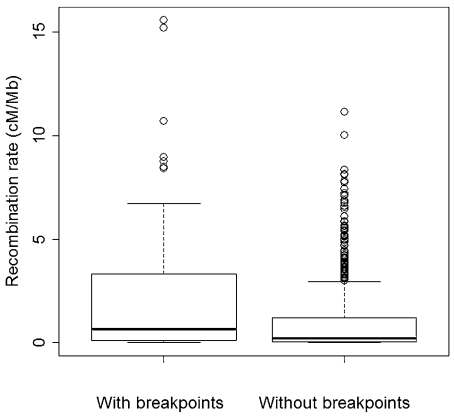

Figure 3. Box-whisker plot representing the recombination rates in 118 1-Mb windows in regions with chromosomal breakpoints (mean ± SD = 2.13 ± 3.03 cM/Mb) and in 643 1-Mb windows without chromosomal breakpoints (1.14 ± 1.95 cM/Mb). The observed difference between the two is statistically significant (Wilcoxon's rank sum test, W = 6376, P = 0.0000183).

Table 2. Recombination rates in 1-Mb nonoverlapping windows in the zebra finch (*Taeniopygia guttata*, TGU) genome that contained either CNVRs only or both CNVRs and chromosomal breakpoints

CNVR ID	Chromosomal breakpoint(s) present?	Recombination rate (cM/Mb)
1	No	0.0448
2	No	0
3	No	12.9356
4	No	n/a
5	No	0.0625
6	Yes	4.5498
7	No	0.6804
8	No	n/a
9	No	2.8987
10	Yes	n/a
11	Yes	n/a
12	No	3.5381
13	Yes	n/a
14	Yes	15.5701
15	No	n/a
16	Yes	0.1806
17	Yes	0.7463
18	Yes	1.3407
19	No	0.5575
20	No	0.9955
21	No	n/a
22	No	n/a
23	No	0.7415
24	No	0.3693
25	No	0.6943
26	No	0
27	No	n/a

n/a, Not available.

Specifically, the conclusions and significant findings of this paper are that:

1. This study and the zebra finch genome paper that it accompanies (Warren et al. 2010) present the first detailed comparative cytogenetic map between two avian species by a combination of in silico and laboratory-based methods.
2. This study is the first to demonstrate the presence of interspecific CNVs across an evolutionary distance of more than 100 Myr (thus providing proof of principle for future studies of the evolutionary significance of CNVs).
3. Our findings demonstrate a highly significant association of both CNV and chromosomal rearrangements with elevated recombination rates. This strongly suggests a role for recombination-based processes in avian genome evolution (and possibly in eukaryotic genomes in general) either as a cause or a consequence of structural genomic rearrangement. Our findings are consistent with results from analyses of mammalian genomes that indicated that NAHR is a major cause of CNVs and chromosomal rearrangements (Stankiewicz and Lupski 2002; Lupski and Stankiewicz 2005; Gu et al. 2008; Hastings et al. 2009).

In addition to the global correlation of structural genomic variation and elevated recombination rates, some specific results of the present study provide further support for the hypothesis that structural genomic variation and recombination are correlated. That is, we observed that seven out of 20 CNVs (35%) for which plausible BLAT results and recombination rates were available were associated with chromosomal rearrangements. This number is higher than expected by chance alone (16%, or 118 out of 761

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

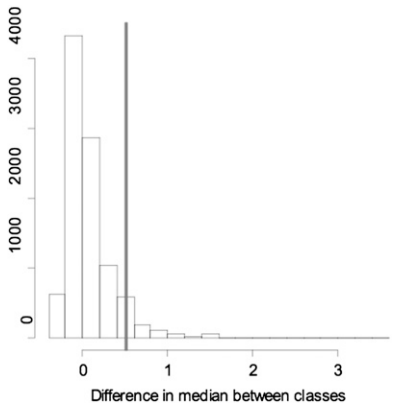

Figure 4. An illustration of the randomization test. The histogram shows the distribution of the difference in median recombination rate between 17 windows drawn at random from the entire sample and the remaining windows. The thick vertical line represents the observed difference in median recombination rate between the 17 windows that contain CNVRs (median = 0.74 cM/Mb) and the remaining 741 windows (median = 0.23 cM/Mb).

1-Mb windows), which points to a common mechanism of origin of both types of structural variation. In addition, the average recombination rate in windows containing both CNVs and chromosomal breakpoints was higher (4.86 cM/Mb) than for windows with either CNVs (2.37 cM/Mb) or breakpoints (1.96 cM/Mb) only. However, the sample size for this class was only $n = 7$, which highlights the need for further studies.

In the present study, recombination rate was measured in nonoverlapping 1-Mb windows. Recent studies in humans, chimpanzees, and mice have shown that recombination rate can also vary substantially on a local (smaller than 1 Mb) scale (McVean et al. 2004; Myers et al. 2005; Ptak et al. 2005; Shifman et al. 2006). Hence, measuring recombination rates based on 1-Mb windows may not fully reflect variation within these windows, which raises the possibility that some structural variants that we observed in 1-Mb windows with apparently low recombination rates may actually occur in small regions with high recombination rates within these windows (and vice versa). We also note that some CNVs and chromosomal rearrangements may also be linked to mechanisms that are not related to recombination, such as non-homologous end-joining (NHEJ), Fork Stalling and Template Switching (FoSTeS), or microhomology-mediated break-induced replication (MMBIR) (Gu et al. 2008; Kim et al. 2008; Arlt et al. 2009; Hastings et al. 2009).

An important point of note is whether the approach used in this study to detect relative CNVs between species is reflective of absolute differences in copy number that could be ascertained by sequence analysis. The definition of a CNV is "a difference in copy number compared to a reference genome" and is, most usually, measured by array CGH. Indeed, using an array CGH approach cross-species (such as in this study) is now reasonably well established in the literature (Redon et al. 2006; Griffin et al. 2008; Skinner et al. 2009). It is certainly the case, however, that differences in sequence divergence might appear as an apparent loss (but not a gain) in the test genome compared to the reference using this approach. For the purposes of this study, however, we would argue first that the aim of our work was to determine whether there are copy number differences between the two species, and thus the question of absolute copy number is not relevant to our primary hypothesis; further sequence-based studies will, ultimately, absolutely determine copy number. Second, we would argue that, in these experiments at least, aberrantly called losses due to sequence divergence were rare due to the fact that we detected more gains than losses. Third, we would argue that we accurately detected whole-chromosome copy number differences of the sex chromosomes, thereby giving us confidence in the ability of our approach to provide a general overview of relative copy number changes between zebra finch and chicken. Thus, while we are in no doubt that results on apparent relative copy number losses (but not gains) may be confounded by sequence divergence, we nonetheless contend that array CGH is a robust means of obtaining a reasonably accurate map of interspecific copy number variation.

In addition to the observed association of structural variation with recombination, our results also paint a more diverse picture of structural changes during avian genome evolution than previous studies (Griffin et al. 2007, 2008; Skinner et al. 2009). In particular, the finding of intrachromosomal rearrangements in almost all chromosomes is in striking contrast to the extraordinary interchromosomal evolutionary conservation of synteny and karyotype structure in avian genomes (Christidis 1990; Rodionov 1997; Griffin et al. 2007). It has been suggested that the low number of interchromosomal rearrangements during avian genome evolution is a consequence of the small amount of interspersed repeats, segmental duplications, and pseudogenes in avian genomes, which provide little opportunity for NAHR (Burt et al. 1999; Burt 2002). However, the detailed comparative maps of the chicken and zebra finch genomes presented here suggest that the organization of bird genomes is more plastic than previously appreciated based on chromosome banding and chromosome painting data. The finding of extensive intrachromosomal rearrangements on a general background of few interchromosomal rearrangements is consistent with genetic mapping studies in the zebra finch, house sparrow (*Passer domesticus*), and Siberian Jay (*Perisoreus infaustus*) (Hale et al. 2008; Stapley et al. 2008; Jaari et al. 2009; Backström et al. 2010). Taken together, these results raise the intriguing question of why overall bird genome structure is so conserved when rearrangements of marker order are so common; in other words, why are interchromosomal rearrangements so much rarer in birds than intrachromosomal rearrangements?

Methods

Detection of chromosomal rearrangements and homologous synteny blocks (HSBs) by whole-chromosome alignments

In order to visualize large-scale intrachromosomal rearrangements, we aligned whole-chromosome sequences of orthologous chicken and zebra finch chromosomes using the program GenAlyzer (Choudhuri et al. 2004) with default settings. For a more detailed analysis of HSBs and chromosomal breakpoints, we retrieved orthologous chicken and zebra finch gene pairs from BioMart (http://www.ensembl.org/biomart/martview/, Ensembl 55 database, *Gallus gallus* genes [WASHUC2] and *Taeniopygia guttata* genes [taeGut3.2.4] data sets) and identified HSBs using the software SyntenyTracker (Donthu et al. 2009) (http://www-app.igb.uiuc.edu/labs/lewin/donthu/Synteny_assign/html/index.html) with default parameters (distance between markers 1 Mb, block size 0 bp, block length 2 Mb, jumping distance 2 Mb, reference genome "zebra finch," target genome "chicken"). We included only genes with orthology type "ortholog_one2one" and with known physical location in both genomes in this analysis. We concatenated overlapping HSBs and treated nested HSBs as independent (i.e., each nested HSB contributed

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

two breakpoints at the end positions of the inserted HSB). Regions between HSBs were considered as chromosomal breakpoints.

Physical mapping using FISH

We used fluorescent in situ hybridization (FISH) to validate a subset of the tentative chromosomal rearrangements suggested by the bioinformatics analyses and to examine conservation of synteny in chicken and zebra finch microchromosomes. Preliminary experiments demonstrated that, probably due to the large evolutionary distance between chicken and zebra finch, cross-species FISH with chicken probes on zebra finch chromosomes did not work sufficiently well to allow for the construction of a comprehensive comparative cytogenetic map. Therefore, we isolated 131 chicken and 131 zebra finch BACs with orthologous sequence content from the Wageningen chicken BAC library (Crooijmans et al. 2000) and the Clemson University Genomics Institute zebra finch BAC library (http://www.genome.clemson.edu/), respectively, and used these BACs to generate two corresponding cytogenetic maps for both species. We determined sequence orthology by aligning chicken sequences retrieved from Ensembl (http://www.ensembl.org/Gallus_gallus/Info/Index, WASHUC2 assembly) against the zebra finch genome using BLASTN (Washington University Genome Sequencing Center BLAST server, http://genomeold.wustl.edu/tools/blast/, running WU-BLAST 2.0 with default settings, database Taeniopygia_guttata-3.2.4 [chromosomes]). Chicken and zebra finch metaphase chromosomes were prepared from lymphocytes and fibroblasts according to standard protocols (Itoh and Arnold 2005; Griffin et al. 2008; Skinner et al. 2009). Isolation of BACs, labeling for single and dual color experiments by nick translation, and FISH followed (Skinner et al. 2009). Briefly, BACs were isolated by midi-preparation and labeled by nick translation with either biotin-dUTP or digoxigenin-dUTP. FISH involved overnight hybridization, stringency washes with 50% formamide, detection with Cy3-strepavidin or FITC anti-digoxigenin, and counterstaining with DAPI. Images were captured using an Olympus BX61 epifluorescence microscope attached to cooled CCD camera and SmartCapture software (Digital Scientific UK). We used the software ImageJ (Abramoff et al. 2004) to measure the position of FISH signals, which was determined as the fractional length from the p terminus, FLpter (Lichter et al. 1990). Supplemental Table 2 summarizes the results of the BLASTN alignments and the BAC hybridizations.

Analysis of relative copy number variation

The detection of relative CNVs in the zebra finch compared to the chicken followed the approach previously used for the analysis of copy number variation in turkey (Griffin et al. 2008) and duck (Skinner et al. 2009). We interrogated the Roche NimbleGen chicken whole-genome tiling array (Catalogue Number/Design Name B3791001-00-01, galGal3 WG CGH; Roche NimbleGen) with whole genomic DNA from one male and one female zebra finch. The Roche NimbleGen array contains 385,000 50-mer oligonucleotides with an average spacing of 2586 bp (source UCSC, build galGal3). We extracted genomic DNA from blood using a DNeasy Animal Blood and Tissue kit (QIAGEN No. 69504); the reference (Red Junglefowl) DNA, from the same animal used in The Chicken Genome Sequencing Project, was kindly provided by Hans Cheng (Michigan State University). Roche NimbleGen performed labeling of genomic DNA by random priming and hybridization to the array. All hybridizations used two dyes per slide (Cy3 and Cy5); Red Junglefowl reference DNA was co-hybridized with zebra finch test DNA.

Roche NimbleGen performed the initial CGH analysis, which proceeded in three stages: normalization, window averaging, and segmentation. After combining the signal intensity and genomic coordinate information, the Cy3 and Cy5 signal intensities were normalized to one another using Qspline normalization (Workman et al. 2002) and prepared for DNA segmentation analysis. This included a window averaging step, where the probes that fall into a defined base pair window size (25 kb) were averaged, using Tukey's biweight mean (Tukey 1960). A new position was assigned to this average, which is the midpoint of the window. Segmentation was also performed on unaveraged data to permit smaller segments than the window size to be detected. The circular binary segmentation algorithm (Olshen et al. 2004) was used to segment the averaged $\log_2$ ratio data. DNA segments were called by attempting to break the segments into subsegments by looking at the *t*-statistic of the means. Permutations (n = 1000) were used to provide the reference distribution. If the resulting *P*-value was below the threshold (default of $P = 0.01$), then a breakpoint was called. A pruning step was used to remove spurious segments, rejecting segments where the standard deviation of the means was not sufficiently different. By default, a cutoff of 1.5 standard deviations was used. CNVs were called for segments in which the $\log_2$ ratio was $>|\pm 0.5|$. Roche NimbleGen provided the results of this analysis in the form of CSV (comma-separated variable) files containing the genomic location of each segment detected by their algorithm. We designed a template in Microsoft Excel to collate these data from both specimens analyzed and further interrogate them. Following the approach of Redon et al. (2006), we called only CNVs containing more than four probes. We combined overlapping CNVs into CNV regions (CNVRs) irrespective of the degree of overlap or whether the change was a gain or loss relative to the Red Junglefowl. It is noteworthy therefore that relative changes in fluorescent intensity detect apparent differences in DNA copy number between test and reference samples (in this case, the two species), not absolute numbers of copies. Moreover, significant sequence variation between the test and the reference may also lead to an apparent relative loss of copy number by this approach. We examined individual CNVRs for association with known or novel genes on the Ensembl database (release 52) (Hubbard et al. 2009). Further information about the array, the experimental design as well as the raw and normalized signal intensities, is available through ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) under the accession number E-TABM-841.

Recombination rate analyses

We obtained sex-averaged recombination rate estimates for 761 1-Mb nonoverlapping windows by comparing genetic and physical postions of ~1400 SNPs evenly distributed along the zebra finch genome (for details, see Backström et al. 2010). To assess if the recombination rate differed between regions with and without chromosomal breakpoints, we partitioned the recombination data into two classes, one with windows containing at least one breakpoint and one with windows without breakpoints, using breakpoint data from the SyntenyTracker analysis (see above). We applied a non-parametric test (Wilcoxon's rank-sum test as implemented in R; http://www.r-project.org/) to assess the level of significance for the difference in recombination rate between these two classes.

Identification of the genomic coordinates of CNVs in the zebra finch genome

We identified the physical positions of CNVs in the zebra finch genome by homology searches with the entire chicken regions containing CNVs as queries in searches against the zebra finch genome. We used the BLAT algorithm (http://www.ensembl.org/Gallus_gallus/), which takes sequence divergence and chromosomal

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

rearrangements into account. This analysis returned best hits on the expected orthologous zebra finch chromosome for 22 out of 27 CNVs; we excluded the remaining five CNVs from the analysis (see Results for details).

Recombination rates and CNVs

Recombination rate data were available for 17 CNVRs with a plausible location in the zebra finch genome according to BLAT alignments. Since the number of CNV-containing windows was much smaller than the total number of windows, we used a randomization test to test for a difference in recombination rates between CNVR-containing windows and all other windows. Using a script in R (http://www.r-project.org/) developed in-house, we generated a distribution (10,000 iterations) of the difference in median recombination rate between 17 windows randomly drawn from the entire sample and the remaining windows and calculated the fraction of expected values higher than the observed (empirical) difference in median to get the *P*-value.

Randomization tests were also applied to evaluate the difference in recombination rates between five windows containing both CNVRs and chromosome breakpoints and windows with either only breakpoints or only CNVRs. In these analyses, five windows were drawn at random to generate the expected distribution of differences in median between classes.

Acknowledgments

We thank Yuichiro Itoh for zebra finch chromosome preparations, Hans Cheng for Red Junglefowl DNA, Zhiguang Wu and Peter Kaiser for chicken blood samples, and the BBSRC for the project grant (BBE010652/1) that funded this study. We also thank The Genome Center, Washington University School of Medicine (specifically The International Zebra Finch Genome Sequencing Consortium) for making the *Taeniopygia guttata* genome sequence available prior to publication.

References

Abramoff MD, Magelhaes PJ, Ram SJ. 2004. Image processing with ImageJ. *Biophotonics Int* **11:** 36–42.

Arlt MF, Mulle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, Glover TW. 2009. Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet* **84:** 339–350.

Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Schneider M, Kempenaers B, Ellegren H. 2010. The recombinational landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* (this issue). doi: 10.1101/gr.101410.109.

Bailey JA, Eichler EE. 2006. Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat Rev Genet* **7:** 552–564.

Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* **96:** 97–112.

Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, et al. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* **402:** 411–413.

Cheng Z, Ventura M, She XW, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437:** 88–93.

Choudhuri JV, Schleiermacher C, Kurtz S, Giegerich R. 2004. GenAlyzer: Interactive visualization of sequence similarities between entire genomes. *Bioinformatics* **20:** 1964–1965.

Christidis L. 1990. *Animal Cytogenetics 4: Chrodata 3 B: Aves*. Gebrüder Borntraeger, Berlin, Germany.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. 2005. Chromosome evolution in eukaryotes: A multi-kingdom perspective. *Trends Genet* **21:** 673–682.

Crooijmans RP, Vrebalov J, Dijkhof RJ, Poel JJ, Groenen MA. 2000. Two-dimensional screening of the Wageningen chicken BAC library. *Mamm Genome* **11:** 360–363.

Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG. 2003. Engineering evolution to study speciation in yeasts. *Nature* **422:** 68–72.

Donthu R, Lewin H, Larkin D. 2009. SyntenyTracker: A tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC Res Notes* **2:** 148. doi: 10.1186/1756-0500-2-148.

Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* **17:** 1266–1277.

Ferguson-Smith MA, Trifonov V. 2007. Mammalian karyotype evolution. *Nat Rev Genet* **8:** 950–962.

Goidts V, Cooper DN, Armengol L, Schempp W, Conroy J, Estivill X, Nowak N, Hameister H, Kehrer-Sawatzki H. 2006. Complex patterns of copy number variation at sites of segmental duplications: An important category of structural variation in the human genome. *Hum Genet* **120:** 270–284.

Griffin DK, Robertson LB, Tempest HG, Skinner BM. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117:** 64–77.

Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryusheva S, Gaginskaya E, Carré W, et al. 2008. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics* **9:** 168. doi: 10.1186/1471-2164-9-168.

Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, Crooijmans RPMA, Besnier F, Lathrop M, Muir WM, Wong GK-S, et al. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res* **19:** 510–519.

Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. *Pathogenetics* **1:** 4. doi: 10.1186/1755-8417-1-4.

Hale MC, Jensen H, Birkhead TR, Burke T, Slate J. 2008. A comparison of synteny and gene order on the homologue of chicken chromosome 7 between two passerine species and between passerines and chicken. *Cytogenet Genome Res* **121:** 120–129.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nat Rev Genet* **10:** 551–564.

Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst* **39:** 21–42.

Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, et al. 2009. Ensembl 2009. *Nucleic Acids Res* **37:** D690–D697.

Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet* **24:** 238–245.

The International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695–716.

Itoh Y, Arnold AP. 2005. Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Res* **13:** 47–56.

Jaari S, Li MH, Merila J. 2009. A first-generation microsatellite-based genetic linkage map of the Siberian jay (*Perisoreus infaustus*): Insights into avian genome evolution. *BMC Genomics* **10:** 1. doi: 10.1186/1471-2164-10-1.

Kim PM, Lam HYK, Urban AE, Korbel JO, Affourtit J, Grubert F, Chen X, Weissman S, Snyder M, Gerstein MB. 2008. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res* **18:** 1865–1874.

Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247:** 64–69.

Lindsay SJ, Khajavi M, Lupski JR, Hurles ME. 2006. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am J Hum Genet* **79:** 890–902.

Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D, Eichler EE. 2003. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13:** 347–357.

Lupski JR, Stankiewicz P. 2005. Genomic disorders: Molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* **1:** e49. doi: 10.1371/journal.pgen.0010049.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461:** 747–753.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA, et al. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457:** 877–881.

McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304:** 581–584.

Downloaded from genome.cshlp.org on June 6, 2013 - Published by Cold Spring Harbor Laboratory Press

Mok BW, Ribacke U, Sherwood E, Wahlgren M. 2008. A highly conserved segmental duplication in the subtelomeres of *Plasmodium falciparum* chromosomes varies in copy number. *Malar J* **7:** 46. doi: 10.1186/1475-2875-7-46.
Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310:** 321–324.
Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* **40:** 1124–1129.
Nair S, Miller B, Barends M, Jaidee A, Patel J, Mayxay M, Newton P, Nosten F, Ferdig MT, Anderson TJC. 2008. Adaptive copy number evolution in malaria parasites. *PLoS Genet* **4:** e1000243. doi: 10.1371/journal.pgen.1000243.
Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. 2009. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res* **19:** 491–499.
Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci* **98:** 12084–12088.
Olshen AB, Venkatraman ES, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5:** 557–572.
Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Caceres AM, Iafrate AJ, Tyler-Smith C, Scherer SW, Eichler EE, et al. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci* **103:** 8006–8011.
Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39:** 1256–1260.
Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, et al. 2008. Copy number variation and evolution in humans and chimpanzees. *Genome Res* **18:** 1698–1710.
Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. *Nat Genet* **37:** 429–434.
Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. 2006. Global variation in copy number in the human genome. *Nature* **444:** 444–454.
Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16:** 351–358.
Rodionov AV. 1997. Evolution of avian chromosomes and linkage groups. *Russ J Genet* **33:** 605–617.
Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, et al. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77:** 78–88.
Sharp AJ, Cheng Z, Eichler EE. 2006. Structural variation of the human genome. *Annu Rev Genom Hum G* **7:** 407–442.
Shaw CJ, Lupski JR. 2004. Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum Mol Genet* **13:** R57–R64.
She X, Cheng Z, Zollner S, Church DM, Eichler EE. 2008. Mouse segmental duplication and copy number variation. *Nat Genet* **40:** 909–914.
Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J. 2006. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol* **4:** e395. doi: 10.1371/journal.pbio.0040395.
Skinner BM, Robertson LBW, Tempest HG, Langley EJ, Ioannou D, Fowler KE, Crooijmans RPMA, Hall AD, Griffin DK, Völker M. 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics* **10:** 357. doi: 10.1186/1471-2164-10-357.
Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends Genet* **18:** 74–82.
Stapley J, Birkhead TR, 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.
Tukey J. 1960. A survey of sampling from contaminated distributions. In *Contributions to probability and statistics* (ed. I Olkin), pp. 448–485. Stanford University Press, Stanford, CA.
Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, Matsumoto N. 2005. Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am J Hum Genet* **76:** 52–67.
Wain LV, Armour JAL, Tobin MD. 2009. Genomic copy number variation, human health, and disease. *Lancet* **374:** 340–350.
Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, et al. 2010. The genome of the zebra finch: Special insights into vocal learning and communication. *Nature* (in press). doi: 10.1038/nature08819.
Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* **3:** research0048.1–research0048.16. doi: 10.1186/gb-2002-3-9-research0048.
Zhang F, Gu WL, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease, and evolution. *Annu Rev Genom Hum G* **10:** 451–481.

Received November 29, 2009; accepted in revised form February 8, 2010.