J Hum Genet (2008) 53:637–643
DOI 10.1007/s10038-008-0292-0

ORIGINAL ARTICLE

Variation of the oxytocin/neurophysin I (*OXT*) gene in four human populations

Yang Xu · Yali Xue · Asan · Allan Daly · Lijie Wu · Chris Tyler-Smith

Received: 16 November 2007 / Accepted: 2 April 2008 / Published online: 20 June 2008

Abstract Oxytocin is a short peptide with multiple functions in human biology and has been implicated in autism. We aimed to determine the normal pattern of variation around the oxytocin gene and resequenced it and its flanking regions in 91 individuals from four HapMap populations and one chimpanzee. We identified 14 single nucleotide polymorphisms (SNPs), all noncoding, including eight that were novel. Population genetic analyses were largely consistent with a neutral evolutionary history, but an Hudson-Kreitman-Aguadé (HKA) test revealed more variation within the human population than expected from the level of chimpanzee–human divergence.

Keywords Oxytocin · *OXT* · Human genetic variation · HapMap populations · HKA test

Introduction

The nine-amino-acid peptide oxytocin is both a hormone and neurotransmitter in mammals, including humans. Its peripheral actions include stimulating cervical dilation and uterine contraction during childbirth, milk release in lactating mothers, and orgasm during sexual activity. In the brain, it influences cognitive, emotional, and social functions, including maternal and sexual behaviors (Caldwell and Young 2006). Particularly striking illustrations of these are the findings that intranasal oxytocin administration increases "mind-reading" ability (Domes et al. 2007) and trust, the latter by specifically increasing an individual's willingness to accept social risks arising through interpersonal interactions (Kosfeld et al. 2005). Such characteristics might have been particularly important during the recent evolution of large complex human societies, so oxytocin regulation might have undergone selection during recent evolution.

In addition, oxytocin has been linked to autism (OMIM 209850) (Carter 2007). This complex phenotype is characterized by poor verbal communication, decreased social interaction, and stereotyped patterns of behavior. It shows high heritability but a complex etiology that is poorly understood, although it includes de novo mutations in a small proportion of autism patients. Among these was a deletion that removed a ~1.1 Mb region containing 27 genes from chromosome 22 (Sebat et al. 2007), including that coding for oxytocin, *OXT*. This was of particular interest because of oxytocin's influence on social interactions noted above (Kosfeld et al. 2005; Domes et al. 2007), because oxytocin plasma levels are decreased in autistic subjects (Modahl et al. 1998), and because intravenous oxytocin has been reported to improve some autistic symptoms, such as the retention of social information (Hollander et al. 2007). If oxytocin level is relevant to autism, could genetic variants that decrease expression increase the chances of developing autism in nondeleted patients when combined with other genetic or environmental factors?

Electronic supplementary material The online version of this article (doi:10.1007/s10038-008-0292-0) contains supplementary material, which is available to authorized users.

Y. Xu · Y. Xue · Asan · A. Daly · C. Tyler-Smith (✉)
The Wellcome Trust Sanger Institute,
Wellcome Trust Genome Campus,
Hinxton, Cambs. CB10 1SA, UK
e-mail: cts@sanger.ac.uk

Y. Xu · L. Wu (✉)
Department of Children's and Adolescent Health,
Public Health College of Harbin Medical University,
157 Baojian Road, Nangang District,
150086 Harbin, People's Republic of China
e-mail: wljhyd@ems.hrbmu.edu.cn

Oxytocin is synthesized as a precursor that also contains a signal peptide and the protein neurophysin I, which are together encoded by the *OXT* gene. The role of neurophysin I is unclear, but it may be important for protection or packaging of oxytocin (Caldwell and Young 2006). We undertook a study of the genetic variation in and around the *OXT* gene in healthy individuals from four populations to discover the full spectrum of variation and assess the selective pressures that have shaped the region during recent human evolution.

Materials and methods

DNA samples

Human DNA samples were from the HapMap populations (The International HapMap Consortium 2005): 23 Yoruba from Ibadan, Nigeria (YRI), 23 Chinese Han from Beijing (CHB), 22 CEPH Utah, USA, residents with ancestry from northern and western Europe (CEU), and 23 Luhya from Webuye, Kenya (LWK). DNAs were purchased from the Coriell Institute for Medical Research (Camden, NJ, USA). In addition, one chimpanzee (*Pan troglodytes*) sample from the European College of Cell Cultures (ECACC) (Salisbury, Wiltshire, UK) was included as an outgroup.

Resequencing and detection of variants

A ~2.2-kb region [Chr20: 2999590–3001789, National Center for Biotechnology Information (NCBI) build 36] was amplified from genomic DNA with the primer pair 1F and 4R (Table 1). The region is highly GC rich, and successful amplification required the addition of dimethyl sulfoxide (DMSO) to the reaction. The 25-µl polymerase chain reaction (PCR) mixture contained 1 × *Taq* buffer (Invitrogen), 200 µM deoxyribonucleotide triphosphate (dNTP), 2 mM magnesium sulfate ($MgSO_4$), 0.4 µM 1F primer, 0.4 µM 4R primer, 1.5 U of Platinum High Fidelity *Taq* DNA polymerase (Invitrogen), 5% DMSO, and 200 ng template DNA. The reaction was initiated by incubating at 94°C for 15 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 3 min, and included a final extension at 68°C for 7 min 30 s. Then, 10 µl PCR product was purified using exonuclease I (0.067 U) and shrimp alkaline phosphatase (0.67 U) for 1 h at 37°C, followed by 15 min at 85°C to denature the enzymes, and then sequenced by the Faculty Small Sequencing Projects Group at The Wellcome Trust Sanger Institute with all four pairs of primers (Table 1) to produce overlapping reads covering both strands.

All potentially polymorphic positions were flagged by the Mutation Surveyor v. 2.0 software (SoftGenetics, State College, PA, USA) and checked manually. Variable positions were compared in overlapping and complementary reads in all individuals, so that most variants were confirmed on both strands in multiple reads. In addition, four blind replicate samples were included in the experiment and showed complete concordance. As a quality check against external data, genotype calls were compared with HapMap calls for the three single nucleotide polymorphisms (SNPs) from this region typed by HapMap (rs2740210, rs2770378, rs2740208) in the 68 common samples (The International HapMap Consortium 2005). Six discrepancies were found (Electronic supplementary material Table 1), mostly corresponding to heterozygote calls in one data set called as homozygotes in the other (overall 1.7% allele difference). Our calls at these positions appeared clear and were used in our analyses. All novel SNPs were confirmed by genotyping using a SNaPshot primer extension protocol (Applied Biosystems; primers in ESM Table 2, results in ESM Fig. 1).

Statistical analyses

Derived alleles were identified from the chimpanzee and macaque sequences (The Chimpanzee Sequencing and Analysis Consortium 2005, Gibbs et al. 2007), and derived allele frequencies were determined by direct counting in the four populations. The following statistics were calculated

Table 1 Primer sequences, polymerase chain reaction (PCR) product sizes and overlap region sizes

Primer name	Primer sequence (5′-3′)	Start	End	Product size (bp)	Overlap (bp)
OXT_1F	TTTGTGACAGCTGTGATAGGAAG	2999590		630	
OXT_1R	GGAGCTCTGTTTAAGAGGTTGGT		3000219		
OXT_2F	GTACTGGGAGGCTGGATAAAGTC	2999959		829	260
OXT_2R	CTTCCGCGCAGCAGATATT		3000787		
OXT_3F	GAACTCCAGGAGCTGAGCG	3000509		670	278
OXT_3R	GACACGAGTCAAGGTAGAGGAGA		3001178		
OXT_4F	ACTCCACCTCTTCCTCCAGAC	3000991		799	187
OXT_4R	TTTATTGCACAGGCAGGAGTAG		3001789		

Table 2 Characteristics of oxytocin/neurophysin I (*OXT*) variants in four HapMap populations

Variant[a]	Position Chr20	Location	dbSNP	Derived allele frequency[b] (%)				
				CEU	CHB	YRI	LWK	
375 G/A	2999924	5′ flanking		0	0	0	2.2	
599 G/A	3000148	5′ flanking		0	0	2.2	0	
1003 C/T	3000552	Intron1		4.5	0	0	0	
1017 G/A	3000566	Intron1	rs6139004	18.2	8.7	8.7	4.3	
1037 G/A	3000586	Intron1		4.5	13.0	4.3	2.2	
1455 C/T	3001004	Intron2		0	0	10.9	6.5	
1641 G/A	3001190	3′ flanking		0	0	4.3	0	
1706 C/A	3001255	3′ flanking	rs2740210	36.4	21.7	10.9	19.6	
1824 G/T	3001373	3′ flanking	rs6115781	0	0	2.2	4.3	
1912 C/G	3001461	3′ flanking		0	2.2	0	0	
1940 C/G	3001489	3′ flanking		0	0	4.3	15.2	
1947 C/G	3001496	3′ flanking	rs2740209	31.8	19.6	8.7	19.6	
1965 G/A	3001514	3′ flanking	rs2770378	65.9	69.6	76.1	67.4	
2085 C/A	3001634	3′ flanking	rs2740208	18.2	13.0	0	0	
312 −/T	2999681	5′ flanking	rs34304890	No data in NCBI				Not found in our study
1404 −/A/G −/A	3000953	Intron2	rs6051569	No data in NCBI				Not found in our study
			rs34097556	No data in NCBI				
1530 A/G	3001079	Exon3	rs17339677	CEU A 2.1 G 97.9				Not found in our study
1748 −/G	3001297	3′ flanking	rs34797740	No data in NCBI				Not found in our study
1998 C/T	3001547	3′ flanking	rs6076471	No data in NCBI				Not found in our study
2113 C/T	3001662	3′ flanking	rs2740207	Not polymorphic in NCBI				Not found in our study

dbSNP Single Nucleotide Polymorphism Database, *NCBI* National Center for Biotechnology Information

[a] Numbered according to GenBank NC_000020

[b] Derived alleles are determined from the chimpanzee sequence, except for variant (1940 C/G), which is from the macaque sequence

using DnaSP 4.0 (Rozas et al. 2003): (1) *Ka*/*Ks*—the ratio of the number of nonsynonymous substitutions per nonsynonymous site (*Ka*) to the number of synonymous substitutions per synonymous site (*Ks*); (2) nucleotide diversity; (3) sequence divergence from chimpanzee; (2) and (3) were also calculated using a sliding window with window length 200 bp and step size 100 bp; (4) allele frequency spectrum summary statistics—Tajima's *D* (Tajima 1989), Fu and Li's *D*, and *F* (Fu and Li 1993) and Fay and Wu's *H* (Fay and Wu 2000); (5) coalescent simulations to evaluate the significance of the results from (4) under a simple neutral model; (2), (4), and (5) were calculated for the whole region and for the *OXT* gene alone; and (6) the Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) was used to compare divergence and diversity at *OXT* with corresponding values from a neutral 498 kb ENCODE region ENr321 (Chr8:118882220-119382220) from chromosome 8q24 (Birney et al. 2007). ENCODE data were only available from the YRI, CHB, and CEU populations, and so the LWK data were omitted from this analysis. As no SNPs were found in the *OXT* coding region, we omitted the coding region of the single gene annotated in ENr321 from this analysis, but similar results were obtained if it was retained. The chimpanzee–human alignment needed for this test was downloaded from Ensembl (http://www.ensembl.org/Homo_sapiens/alignsliceview?c=8:119132220.5;w=500000;align=opt_align_228) and the divergence calculated by DnaSP. The population differentiation at individual sites was measured using F_{ST} (Schneider et al. 2000). Linkage disequilibrium (LD) was investigated using Haploview 3.32 (Barrett et al. 2005). Haplotypes were inferred from the genotype data using PHASE 2.1 (Stephens et al. 2001), and a median-joining network was constructed from the resulting haplotypes using NETW4.1.1.2 (Bandelt et al. 1999) (http://www.fluxus-engineering.com/sharenet.htm).

Results

We first evaluated the evolutionary pressures acting on the *OXT* coding region over the last few million years by examining the *Ka*/*Ks* ratio calculated from the human, chimpanzee (The Chimpanzee Sequencing and Analysis Consortium 2005), and rhesus macaque (Gibbs et al. 2007) reference sequences. This value will be dominated by the

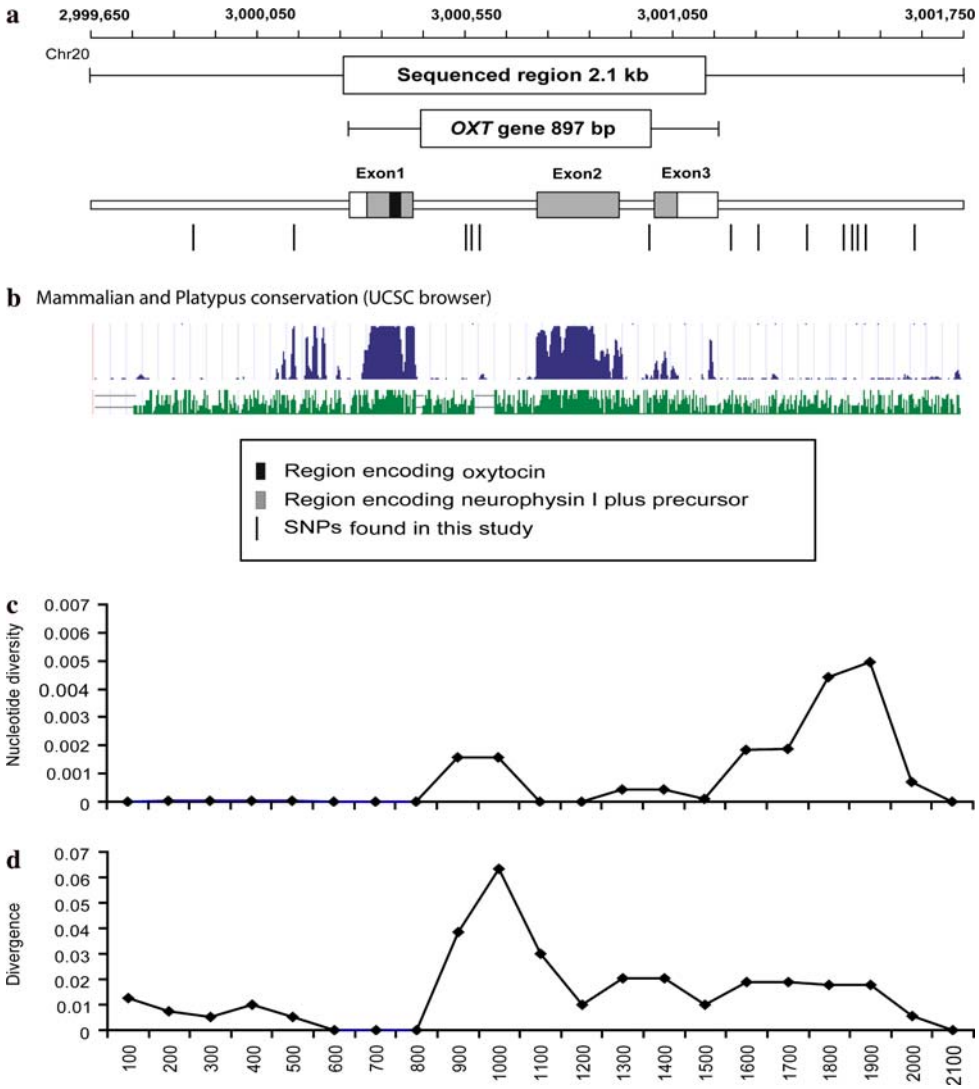

Fig. 1 Oxytocin/neurophysin I (*OXT*) gene structure and variation. **a** Location of regions used for analysis, exons, and coding regions, and single nucleotide polymorphisms (SNPs) discovered. **b** Conservation, illustrated by the "Mammal Cons" and "Platypus" tracks from the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTracks?org=human). **c** Distribution of nucleotide diversity within humans, and **d** divergence between human and chimpanzee along the sequence using a sliding window with length 200 bp and step size 100 bp

neurophysin I sequence because of its length and was 0.51. This is higher than the genomewide average of 0.25 (Gibbs et al. 2007) but lower than the neutral value of 1. It therefore indicates constrained evolutionary change, although with less constraint than the average protein, and supports the idea that the amino acid sequence of neurophysin I is functionally important.

We then determined the sequence of a 2.1-kb region encompassing the gene (Chr20: 2999650-3001750, Fig. 1) in 91 individuals from four populations and one chimpanzee. We found 14 variable positions in humans, all of which were base substitutions and lay in the flanking or intronic regions; none were found in the coding region (Table 2; ESM Table 3). Six of the SNPs were present in the Single Nucleotide Polymorphism Database (dbSNP), but the other eight were novel, and all of these were confirmed by genotyping (ESM Fig. 1). As expected, the novel SNPs were lower in frequency than the known ones but included two present at >10% frequency in individual population samples. Six dbSNP entries were not found in our sample: one of these was rare, and the others have not been reported in population surveys so may either represent other rare variants or erroneous database records (Table 2). The average nucleotide diversity for the region was 8.6×10^{-4} and was slightly lower for the *OXT* gene alone (4.5×10^{-4}) (Table 3), but both these values were within the normal range (e.g., Akey et al. 2004). Strikingly, both nucleotide diversity and divergence from chimpanzee varied across the region, being low upstream of the gene and in the exons but higher in intron 1 and downstream of the gene (Fig. 1). On this time scale, the 5' region was almost as constrained as the coding region.

We next investigated whether the patterns of variation within the four populations were consistent with a neutral evolutionary history or indicated the action of positive or balancing selection. The allele frequency spectrum statistics examined (Tajima's *D*, Fu and Li's *D*, Fu and Li's *F*, and Fay and Wu's *H*) can show unusually low values in

Table 3 Oxytocin/neurophysin I (*OXT*) gene summary statistics

Population	Sample size (chromosomes)	Sample characteristics				Allele-frequency distribution tests							
		Polymorphic sites		Nucleotide diversity ($\times 10^{-4}$)		Tajima's *D*		Fu and Li's *D*		Fu and Li's *F*		Fay and Wu's *H* (P^c)	
		OXT gene[a]	Whole region[b]	*OXT* gene[a]	Whole region[b]	*OXT* gene[a]	Whole region[b]	*OXT* gene[a]	Whole region[b]	*OXT* gene[a]	Whole region[b]	*OXT* gene[a]	Whole region[b]
CEU	44	3	7	5.4	10.3	−0.641	0.936	0.908	1.289	0.521	1.393	0.406 (0.344)	0.653 (0.610)
CHB	46	2	7	4.4	8.4	−0.244	0.303	0.752	0.478	0.537	0.500	0.344 (0.496)	0.522 (0.494)
YRI	46	3	10	5.0	6.8	−0.732	−1.073	0.903	0.033	0.483	−0.356	0.402 (0.337)	0.077 (0.337)
LWK	46	3	9	2.8	8.3	−1.326	−0.430	−0.420	−0.133	−0.810	−0.312	0.240 (0.123)	0.371 (0.441)
Total	182	4	14	4.5	8.6	−0.762	−0.639	0.894	−0.504	0.409	−0.644	0.369 (0.325)	0.519 (0.457)

[a] *OXT* gene: 717–1613 fragment (ch20:3000266-3001162)

[b] Whole region: 100–2200 fragment (ch20:2999650-3001750)

[c] *P* value

response to positive selection or high values in response to balancing selection but were all consistent with neutrality, both in individual populations and in the worldwide sample (Table 3). The degree of population differentiation (F_{ST} value), which can be high if positive selection has acted differentially on populations, was not observed to be unusually high, ranging from 0 to 0.17 (ESM Table 4). Eighteen inferred haplotypes were present, and LD was extensive ($D' = 1$) through most of the region (ESM Fig. 2). A median-joining network linking the haplotypes was therefore constructed and showed a compact structure with limited reticulation, as expected from the LD pattern (Fig. 2). The three most common haplotypes were all present in all populations, consistent the lack of population differentiation shown by individual SNPs.

Finally, we assessed whether the level of variation in humans was as expected from the amount of divergence from chimpanzee, as it would be under a strictly neutral model of evolution. For this, we used an HKA test in which *OXT* was compared with other neutral resequenced regions. In order to provide power, we used ~500 kb resequenced by the ENCODE project (Birney et al. 2007) in three of the same populations. More variation was found at *OXT* than expected ($P = 0.024$; Table 4), and in view of the large comparative region analyzed, this is a robust result.

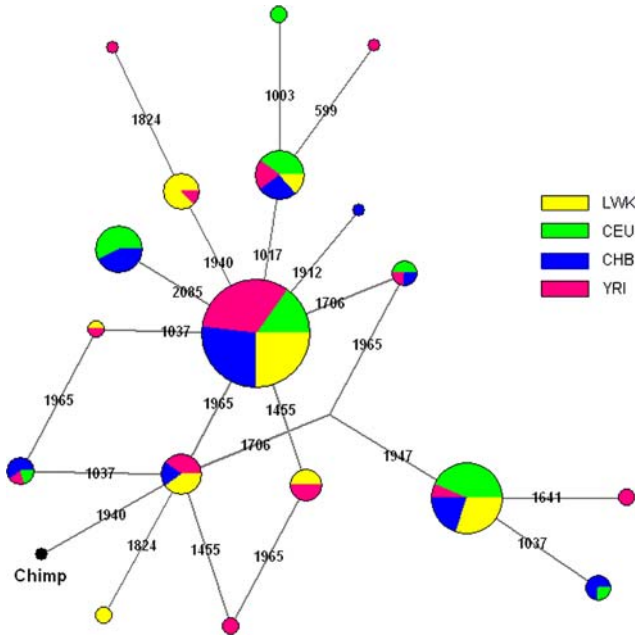

Fig. 2 Median-joining network of oxytocin/neurophysin I (*OXT*) haplotypes. *Circles* represent haplotypes and *lines* the mutational steps that distinguish them. *Numbers* on the lines refer to single nucleotide polymosrphism (SNP) positions in Table 2. *Circle area* is proportional to haplotype frequency, and *circles* are color coded according to population

Discussion

We carried out a thorough investigation of the genetic variation in and around the *OXT* gene. We were particularly interested in two aspects of this: whether or not variants likely to influence function were present, and whether or not the region showed a neutral evolutionary history.

Functional variants are located most obviously in the protein-coding regions of a gene, but evolutionary analyses show that many noncoding sequences are conserved to the same extent as coding sequences and thus are also likely to be functionally important. No *OXT* coding variants were found, but two novel low-frequency SNPs were discovered in the 5′ flanking region in African populations, one of which (at 3,000,148) lies in the most highly conserved segment outside the exons (Fig. 1b). Comparison with a

Table 4 Results of the Hudson-Kreitman-Aguadé (HKA) test

	496.3 kb in 8q24	2.1 kb in 20p13
Intraspecific polymorphism data		
Segregating sites (obs)	1726	13
Segregating sites (exp)	1731.99	7.01
Total no. of sites	499678	2100
Sample size	96	136
Interspecific divergence		
No. differences (observed)	6317.00	18.00
No. differences (expected)	6311.01	23.99

$\chi 2 = 5.096$, $P = 0.024$

Table 5 Comparison of human and platypus sequence near the oxytocin/neurophysin I (*OXT*) gene

Sequence position	Size (bp)	Matched (bp)	Unmatched (bp)	% matched
5′ flanking sequence	616	281	335	46
OXT gene sequence	897	485	412	54
3′ flanking sequence	588	256	332	44
Whole sequence	2101	1022	1079	49

species such as platypus shows that the downstream region is overall as conserved as the upstream (44% downstream, 46% upstream; Table 5), although this conservation is broadly distributed and lacks highly conserved segments (Fig. 1b). It is therefore possible that SNPs in one or both regions might contribute to variation in the expression of the gene.

Most of the evolutionary tests suggested a neutral evolutionary history of the region, ruling out strong recent positive or balancing selection. The HKA test, however, revealed a larger number of variants than expected when the evolutionary divergence rate was taken into account. This finding is in one respect conservative, because the evolutionary divergence may be overestimated when a single chimpanzee is used for comparison. The result can be viewed in two ways: It could be argued that as we applied five tests (Tables 3, 4), a strict Bonferroni correction would require a P value of 0.01 for significance, and thus our observed P value of 0.024 is not significant; or the excess variants are within the range expected by chance. However, a Bonferroni correction is probably too conservative in this case, because the tests in Table 3 all examine related aspects of the allele frequency spectrum and so are not independent. According to this view, the excess of variants is above that expected by chance and, to speculate further, increased diversity of 3′ regulatory elements might have been selected for, perhaps leading to a range of *OXT* expression levels in the population.

It is now possible to design in vitro experiments to study the regulation of *OXT* expression and association studies to relate its variation to phenotypes of interest, such as plasma oxytocin level or autism. It is striking that LD in the 100-kb region surrounding *OXT* is very low in all HapMap populations and is even incomplete for position 3,001,514 within the short region resequenced, so tagging would be inefficient and all SNPs need to be identified and genotyped directly in association studies. The new SNPs discovered in this study provide additional material for such projects.

Acknowledgments We thank the Sanger sequencing team for generating the sequence data. Y. Xu and LW were supported by the National Natural Science Foundation of China (grant no. 30771817) and by Harbin Medical University (No. HCXB2008001); Y. Xue, AD, and CTS were supported by The Wellcome Trust.

References

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G,

Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816

Caldwell HK, Young WS III (2006) Oxytocin and Vasopressin: genetics, behavioral implications. In: Lim R (ed) Handbook of neurochemistry and molecular neurobiology. 3rd edn. Springer, Berlin

Carter CS (2007) Sex differences in oxytocin and vasopressin: implications for autism spectrum disorders? Behav Brain Res 176:170–186

Domes G, Heinrichs M, Michel A, Berger C, Herpertz SC (2007) Oxytocin improves "mind-reading" in humans. Biol Psychiatry 61:731–733

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csuros M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'Brien W E, Prufer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Karolchik D, Kern AD, Kuhn RM, Smith KE, Zwieg AS (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234

Hollander E, Bartz J, Chaplin W, Phillips A, Sumner J, Soorya L, Anagnostou E, Wasserman S (2007) Oxytocin increases retention of social cognition in autism. Biol Psychiatry 61:498–503

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. Nature 435:673–676

Modahl C, Green L, Fein D, Morris M, Waterhouse L, Feinstein C, Levin H (1998) Plasma oxytocin levels in autistic children. Biol Psychiatry 43:270–277

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Schneider S, Roessli D and Excoffier L (2000). Arelquin: a software for population genetics data analysis, Genetics and Biometry Lab, Department of Anthropology, University of Geneva

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimaki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M (2007) Strong association of de novo copy number mutations with autism. Science 316:445–449

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320