Doctoral Thesis

博士論文

Genetic variation of olfactory receptor multigene family in humans

（人類集団における嗅覚受容体多重遺伝子族の遺伝的多様性）

AKHTAR MUHAMMAD SHOAIB

アクター モハマッド シュエブ

Doctoral Thesis

博士論文

Genetic variation of olfactory receptor multigene family in humans

（人類集団における嗅覚受容体多重遺伝子族の遺伝的多様性）

AKHTAR MUHAMMAD SHOAIB

アクター モハマッド シュエブ

Graduate School of Frontier Sciences

The University of Tokyo

Acknowledgements

I am extremely obligate and thankful to my research supervisor Dr. Shoji Kawamura whose sympathetic attitude, inspiring guidance, constructive criticism, encouragement, and co-operation always serve as a beacon of light during this research work. I am grateful to my collaborators Dr. Hiroki Oota (The University of Tokyo), Dr. Hajime Ishida (University of Ryukyus), Dr. Yoshihito Niimura (University of Miyazaki), Dr. Kazushige Touhara (The University of Tokyo) and Dr. Amanda D. Melin (University of Calgary, Canada) for their scientific input and technical support. I am thankful to Dr. Kazuhiro Nakayama, Dr. Yuka Matsushita, Dr. Mari Nishikawa, Mr. Ryuichi Ashino, Ms. Min Hou, Mr. Shu Sun, Ms. Qinyuan Ji and all members of Laboratory of Evolutionary Anthropology for their moral support.

Moreover, I am thankful to The University of Tokyo International Student Support for award of three years fellowship. I am also grateful to International Liaison Office of Graduate School of Frontier Sciences and Kashiwa International Office of The University of Tokyo for their support in everyday life.

I am extremely thankful to my parents and siblings for their continuous support and cooperation. I am also thankful to my all friends and colleagues at The University of Tokyo.

Last but not least, prayers and peace upon Prophet Muhammad whose teachings and practices always make me strong spiritually and keep me motivated.

Akhtar Muhammad Shoaib

List of Abbreviations

1000G	1000 Genomes Project
1KG	1000 Genomes Project
1KG WES	1000 Genomes Project Whole Exome Sequence
1KG WGS	1000 Genomes Project Whole Genome Sequence
ADG	Adygei population
AET	Aeta population
AGT	Agta population
ANU	Ainu population
BAM	Binary Alignment Map
BPG	Biaka Pygmy population
BQSR	Base Quality Score Recalibration
BTK	Batak population
BWA	Burrows-Wheeler Alligner
CHG	Chagga population
CNV	Copy Number Variation
dbSNP	Single Nucleotide Polymorphism Database
DNE	Dane population
FASTA	FAST-All (a genomic sequence format)

F_{ST}	Fixation Index
GATK	Genome Analysis Toolkit
HAS	Hausa population
hg19	Human Reference Genome Assembly build 19
hg38	Human Reference Genome Assembly build 38
HNS	Honshu population
HORDE	Human Olfactory Receptor Database
INDEL	Insertion Deletion
IRS	Irish population
MEM	Maximal Exact Match
MNB	Manobo population
MNW	Mamanwa population
MPG	Mbuti Pygmy population
NGS	Next Generation Sequencing
NR	Neutral Reference
N_{ST}	Coefficient of Nucleotide Variation
nt	Nucleotide
OR	Olfactory Receptor
ORF	Open Reading Frame

PCA	Principal Component Analysis
RUS	Russian population
RYK	Ryukyu population
SAM	Source Alignment Map
SNP	Single Nucleotide Polymorphism
TC	Target Capture
TLG	Tagalog population
VCF	Variant Call Format
VQSR	Variant Quality Score Recalibration
VSN	Visayan population
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Table of Contents

Acknowledgements i

List of Abbreviations ii

Table of Contents v

List of Appendices viii

List of Tables ix

List of Figures x

CHAPTER 1 1

General Introduction with Materials and Methods 1

1.1. Olfaction 2
1.2. Olfactory Receptors 2
1.3. OR Repertoire 3
1.4. Challenges to OR Sequencing 6
1.5. Targeted Capture 6
1.6. The Objective of This Study 7

CHAPTER 2 9

Olfactory Receptor Gene Retrieval Using Targeted Capture Sequencing for Diverse Human Populations 9

2.1. Introduction 10
2.2. Materials and Methods 11
2.2.1. Study Populations 11
2.2.2. Genomic DNA Samples 14
2.2.3. OR Probe Design 16
2.2.4. Neutral Reference Probe Design 18
2.2.5. Targeted Capture & NGS 19
2.2.6. Bioinformatic Analysis 20
2.2.7. Statistical Analysis 23

2.3. Results 23
2.3.1. Sequence Retrieval 23
2.3.2. Sequencing Depth 24
2.3.3. Variant Sites 26
2.4. Discussion 27
CHAPTER 3 30
Segregating Pseudogene Polymorphism and Copy Number Variation of Olfactory Receptor Genes in Diverse Human populations 30
3.1. Introduction 31
3.2. Materials and Methods 32
3.2.1. Study Individuals 32
3.2.2. Sequencing Data 32
3.2.3. Intact/Disrupted Gene Allele Composition 32
3.2.4. CNV Analysis 33
3.3. Results 34
3.3.1. Disruption of OR Genes 34
3.3.2. Copy Number Variation 35
3.4. Discussion 37
CHAPTER 4 41
Population Differentiation of Olfactory Receptor Gene Repertoire in Humans 41
4.1. Introduction 42
4.2. Materials and Methods 43
4.2.1. Materials 43
4.2.2. Population Differentiation 43
4.2.3. N_{ST} Analysis for Population Differentiation 45
4.2.4. Tajima's *D* Analysis 46
4.3. Results 48
4.3.1. SNP-Based Inter-Individual Diversity of OR Genes and Neutral References by PCA 48
4.3.2. N_{ST} Population Differentiation Analysis 49
4.3.3. Tajima's D Analysis 49
4.4. Discussion 50
CONCLUSIONS 54

TABLES 55

FIGURES 61

REFERENCES 95

APPENDICES 115

List of Appendices

Appendix I…….115

Appendix II……....120

Appendix III……...154

Appendix IV…….…..158

Appendix V…….170

Appendix VI……180

Appendix VII……..186

List of Tables

Table 1: Number of individuals of each population 55

Table 2: Number of variant sites and SNP density among study individuals.....56

Table 3: N_{ST} genetic distance measure between study populations 57

Table 4: d_{xy} genetic distance measure between study populations.....58

Table 5: d_A genetic distance measure between study populations.....59

Table 6: Featured OR genes in this study 60

List of Figures

Figure 1: Anatomical representation of human olfactory system. 61

Figure 2: Distribution of variants among structural categories as present in hORMdb. 62

Figure 3: Exemplary schematic illustration of target capture probes of 120 nucleotides each with 60 nucleotides overlapped. 63

Figure 4: Flowchart of NGS Reads Analysis.....64

Figure 5: Retrieval frequency of three Pantro OR genes among study populations.....65

Figure 6: Box plots of sequencing depths for all data throughout genes/regions and throughout carrier individuals in each gene/region category.....66

Figure 7: Comparison of mean depths among targeted captures OR-A, NR-A, OR-X, NR-X and OR-P, the whole-genome sequencing in the 1000 Genomes project (1K WGS) and the whole-exome sequencing in the 1000 Genomes project (1K WES).....67

Figure 8: Box plots of sequencing depths averaged throughout carrier individuals for each gene/region in each gene/region category.....68

Figure 9: Box plots of sequencing depths averaged throughout genes/regions in each gene/region category for each carrier individual 69

Figure 10: Mean depth of autosomal ORs and neutral references per carrier plotted by population.. 70

Figure 11: Frequencies of disrupted alleles of the OR segregating pseudogenes in the 401 study individuals. 71

Figure 12: Disrupted Allele frequencies across study individuals of three Japanese populations.. 72

Figure 13: Disrupted Allele frequencies across study individuals of four negrito populations of Philippines. 73

Figure 14: Disrupted Allele frequencies across study individuals of four agricultural populations of Philippines.....74
Figure 15: Disrupted Allele frequencies across study individuals of four European study populations.....75
Figure 16: Disrupted Allele frequencies across study individuals of four African study populations.....76
Figure 17: Distribution of number of homozygous intact OR genes among each study population. 77
Figure 18: The deletion, duplication and triplication allele frequencies of autosomal OR genes. 79
Figure 19: CNV (deletion, duplication, and triplication) allele frequencies in three Japanese study populations. 80
Figure 20: CNV (deletion, duplication, and triplication) allele frequencies in four negrito study populations of Philippines.....81
Figure 21: CNV (deletion, duplication, and triplication) allele frequencies in three agricultural non-negrito study populations of Philippines.....82
Figure 22: CNV (deletion, duplication, and triplication) allele frequencies in four European study populations. 83
Figure 23: CNV (deletion, duplication, and triplication) allele frequencies in four African study populations. 84
Figure 24: QQ plot to see distribution of Tajima's D among ORs and neutral references. 85
Figure 25: Population differentiation of ORs and neutral references. 86
Figure 26: N_{ST} genetic differentiation as a quantitative measure of population differentiation. 87
Figure 27: d_{xy} distance matrix as a quantitative measure of population differentiation. 88

Figure 28: d_A distance matrix as a quantitative measure of population differentiation. 89

Figure 29: Distribution of N_{ST} values per gene/region among autosomal hg38 – intact ORs and neutral references. 90

Figure 30: Tajima's D of hg38 OR intact and pseudogenes and neutral references examined for all 401 study individuals. 91

Figure 31: Tajima's D of intact ORs, disrupted ORs and neutral references. 92

Figure 32: Tajima's D distribution of OR intact, pseudogenes and neutral references in Filipino study populations.....93

Figure 33: Tajima's D distribution of OR intact, pseudogenes and neutral references in European and African study populations. 94

CHAPTER 1

General Introduction with Materials and Methods

Senses are necessary for everyday life of all living organisms (Geldard et al., 1953, Serres, 2008). Humans are known to have five basic senses including sight (vision), hearing (audition), smell (olfaction), taste (gustation) and touch (mechanoreception). Senses of vision, audition and mechanoreception perceive physical stimuli and known as physical senses. Senses of olfaction and gustation perceive chemical stimuli and known as chemical senses. Both chemical senses, gustation and olfaction, are regulated by G-protein coupled receptor super family. A seven transmembrane structure is hallmark of G-protein coupled receptors (Buck and Axel, 1991, Zhang and Firestein, 2002). Humans are generally considered to be visually oriented animals and the importance of their olfactory sense is not widely appreciated. However, recent studies have challenged the view, revealing significant roles of olfaction in daily life and communication (Hoover, 2010, McGann, 2017, Bushdid et al., 2014, Roberts et al., 2020).

1.1. Olfaction

Olfaction is initiated by olfactory epithelium present in nasal cavity. Olfactory epithelium contains olfactory sensory neurons extending to higher structures of brain as shown in Figure 1 (Buck and Axel, 1991). These olfactory sensory neurons work in combinatorial mode making a many-to-many relationship and can sense more than a trillion odors (Buck and Axel, 1991, Bushdid et al., 2014).

1.2. Olfactory Receptors

Each olfactory sensory neuron is encoded by an olfactory receptor (OR) gene. OR genes comprise the largest gene family in vertebrates. They belong to the G-protein coupled receptor (GPCR) super gene family and have a seven-transmembrane structure (Buck and

Axel, 1991, Zhang and Firestein, 2002). OR genes are single-coding-exon genes lacking introns with a typical size of ~930 bp. Data on the diversity of OR genes in human populations is accumulating, with records of prevalent single-nucleotide polymorphisms (SNP), short insertions and deletions (indels), intact (functional)/disrupted (pseudogene) polymorphisms, and copy number variations (CNVs), that have been associated with olfactory perceptual variations (Menashe et al., 2002, Gilad and Lancet, 2003, Menashe et al., 2007, Keller et al., 2007a, Hasin et al., 2008, Keller et al., 2012, Jaeger et al., 2013b, McRae et al., 2013, Wooding, 2013, Hoover et al., 2015, Secundo et al., 2015, Majid and Kruspe, 2018, Olofsson and Wilson, 2018, Trimmer et al., 2019, Concas et al., 2021).

Until today, 119,069 natural variants including SNPs and indels are reported in human olfactory receptor mutation database (hORMdb) (Jimenez et al., 2021). These variants correspond to 378 OR genes belonging to 17 OR families. Highest of these variants are amino acid altering (non-synonymous) point mutations ("missense" mutations in Figure 2) (64%) and thus can result in olfactory function variation. On other hand, only 25% of variants are reported to be synonymous. Distribution of hORMdb variants is shown in Figure 2. hORMdb contains data from seven subcontinental populations of Europe, Asia and Africa. Among these data, only 2182 variant alleles represent global allele frequency >1%, in contrast >95% of variant alleles remain population specific and low in allele frequency. These variant data suggest that there are many population specific variations in olfactory receptors.

1.3. OR Repertoire

OR multigene family is evolved from a single copy ancestral gene by duplications (Yohe et al., 2020). Duplicated gene are termed as paralogous to each other. This phenomenon raises

variation in OR gene repertoire. As per human reference genome hg38 assembly, there are 840 OR genes, more than of half of them (442) are pseudogenized and 398 are intact (Niimura et al., 2018). Frequency of pseudogenes in OR family is higher than any other gene family (Rouquier et al., 2000, MacArthur et al., 2012). Thus, intact/disrupted composition could be highly variable and give rise to high variation in olfactory perception among humans.

Many OR genes are also reported to be "segregating pseudogenes" i.e., polymorphic in terms of having disrupted forms as allele in addition to intact allele. The human olfactory data explorer (HORDE), a database of human olfactory genes by Weizmann Institute of Science, reports 61 segregating pseudogenes among humans (Olender et al., 2013). In addition to intact/disrupted composition, many OR genes show CNVs. The information on nasal gene expression level and non-nasal expression of OR genes and the relation of such expression data with the intact/disrupted polymorphism and CNVs would be important to understand the variation of olfactory ability. However, invasive nature and technical difficulties of such studies on anatomically internal location of olfactory sensory neurons has been an obstacle.

Nevertheless, gene expression of 273 human ORs had been observed in 26 human autopsies (Verbeurgt et al., 2014). 384 OR genes were targeted in total and patterns of expression were varied across autopsies. Only 90 ORs were expressed among all autopsies, the rest 183 were variable among each autopsy (Verbeurgt et al., 2014). In addition to nasal expression, ORs are also found to express in various other organs including brain, heart, carotid body, kidneys, ovary, and testis (Chang et al., 2015, Carithers and Moore, 2015, Lonsdale et al., 2013, Pluznick et al., 2009, Shepard, 2021).

Ligand information of ORs is also important to understand the functional difference and variation of ORs. However, for many of ORs, ligands remain unassigned due to technical difficulty of the successful heterologous expression of many OR genes in cultured cells (Saito et al. 2009). Difficulty of revealing odor-OR coding also stems from the hugeness of OR gene family size and the many-to-many correspondence between odors and ORs (DeMaria and Ngai, 2010). Nevertheless, ligand information has been reported for about 90 ORs (Appendix I) (Verbeurgt et al., 2014, Trimmer et al., 2019, McRae et al., 2012, McRae et al., 2013, Jaeger et al., 2013b, Jaeger et al., 2013a). Most of these ORs are also known for their nasal expression.

Humans are diverse in living environment, historical subsistence, and culture, which could have influenced the OR repertoire. This could be the reason why human OR repertoire (intact and pseudogene composition) reported by different research groups is different (Hughes et al., 2014, Niimura et al., 2018, Olender et al., 2013, Go and Niimura, 2008). Previous studies reported signatures of natural selection in OR multigene family (Gilad and Lancet, 2003) and OR repertoire of each population could has been shaped by local adaptations to their living environment and culture (Nozawa et al., 2007). There are also reports of influence of historical subsistence on OR repertoire (Majid and Kruspe, 2018, Majid, 2021).

Building on these previous studies, I focus on the human populations across the globe to study OR gene family variation at the nucleotide sequence level. This is the age of genomics, and many genomics data are coming out day by day, but many anthropologically interesting populations remain unstudied even in big genomic projects like 1000 Genomes Project (Consortium, 2012, Consortium, 2015, Consortium, 2005, Consortium, 2007).

1.4. Challenges to OR Sequencing

Humans have ~400 intact and ~440 disrupted OR genes in the genome (Glusman et al., 2001, Hasin et al., 2008, Niimura et al., 2018). Because duplicated genes are similar in DNA sequence and creating assemblies is inherently problematic, many multigene families assemble poorly with low coverage in whole-genome sequence (WGS) databases using short-read massive parallel ("next-generation") sequencing (NGS) (Sims et al., 2014, Yohe et al., 2020). Although the long-read sequencing would be an option for sequencing a multigene family in a genomic region (Larsen et al., 2014, Eaton et al., 2021), combined use of short-read NGS is still desired to increase sequencing accuracy. However, its application to ecological, evolutionary, or population studies is often limited due to difficulty of collecting sufficient amount of high-molecular-weight genomic DNA samples. Since OR genes represent ~1% of genomic coding sequence and only ~0.0125 % of the whole genome in mammals (Lane et al., 2001, Mainland et al., 2013), application of WGS approach to sequence OR genes would also be labor- and cost-ineffective.

1.5. Targeted Capture

Targeted capture is a strategy to enrich subsets of the genome by hybridizing nucleotide probes designed from reference sequences with fragmented genomic DNA of study subjects. The DNA, now enriched in the regions of interest, is sent for subsequent sequencing. This strategy has emerged as a powerful alternative to WGS (Jones and Good, 2016). The growing use of targeted capture with short-read NGS demonstrates its potential to address a range of research questions (Schweizer et al., 2016). Targeted capture has been successfully applied to sequence OR genes with high coverage and to identify variation with high sensitivity and specificity (Mainland et al., 2013, Yohe et al., 2020).

I employed the targeted capture using probes designed mainly from the human reference genome hg38 and conducted the short-read NGS. The hg38 sequence represents a haploid genome sequence at each nucleotide site from a source individual for the genome region. In addition to the annotated OR genes in hg38, I also included non-annotated OR sequences and nearly intact OR gene sequences in hg38 and the OR gene sequences absent in hg38 but present in the chimpanzee reference genome Pantro3.0 for the probe design to enable a comprehensive survey that would capture OR genes possibly missing in hg38 due to genetic polymorphism present in the human population. In order to proxy neutral variation, I additionally captured single-copy and non-protein-coding genome regions as "neutral" control references in the same genomic DNA samples to evaluate genetic diversity of the OR gene family and the neutrality of the variation in the human populations.

1.6. The Objective of This Study

The objective of this study was to find patterns of OR repertoire variation among Asian, European and African populations. In my research, I studied OR repertoire of 18 populations of Asian-, European- and African-origins comprising 401 individuals. Most of these populations remain underreported but represents diverse anthropological background.

In the present study I test whether the targeted capture by probes with diversity-oriented design is more effective than the whole-genome approach to retrieve OR genes and achieve high-depth sequencing. I then examine whether there is any difference of the OR gene composition between this study and the reference human genome data and how novel the identified variation in this study is. I also test whether the application of the same methods to neutral references is effective to reveal natural selection on the OR gene family. I

evaluate whether the variation is associated with difference of genomic ancestry over the globe.

CHAPTER 2

Olfactory Receptor Gene Retrieval Using Targeted Capture Sequencing for Diverse Human Populations

2.1. Introduction

Targeted capture sequencing has revolutionized field of molecular genetics. Comparing with the whole-genome sequencing, key features of targeted capture sequencing includes focusing the sequencing effort on genes/regions of interest, a smaller data volume easy to handle, cheaper in cost, reduced burden to process sequencing reads, fast processing, and higher sequencing depth (Grossi et al., 2020, Amadori et al., 2020, Iadarola et al., 2020, Prabhu and Pe'Er, 2009, Jones and Good, 2016). These wide range of targeted capture sequencing merits have widened its applications from medical applications to evolutionary biology (Jin et al., 2021, Gil-Varea et al., 2020, Liu et al., 2021, Jones and Good, 2016). Due to these key benefits, number of publications utilizing targeted capture sequencing have increased ~1000 times in last 10 years in PubMed (Sayers et al., 2021).

This chapter describes features of OR gene sequences retrieved by the targeted capture sequencing. As I described in Chapter 1 sequencing multigene families is a challenge due to presence of highly similar paralogous genes (Yohe et al., 2020). Thus, most of the publicly available WGS data have poor efficiency in assembling multigene families. I conducted targeted capture with diversity-oriented design of probes to retrieve OR gene sequences with high confidence.

I targeted 501 OR genes from human reference genome hg38 assembly and 53 chimpanzee OR genes for which there are no orthologous genes in hg38. Orthologous genes are the genes separated by speciation in the past, while paralogous genes are the genes separated by gene duplication in the past (Jensen, 2001). These 501 ORs include intact (regarded as functional) and disrupted genes (pseudogenes) as described earlier in Chapter 1. In addition to OR genes, 85 neutral references were also targeted. Thus 554 ORs (501 hg38 and 53 chimpanzee) and 85 neutral references were targeted, captured, and sequenced.

In addition to the targeted capture design, which is described in detail in chapter 1, appropriate processing of sequencing reads is also a key to retrieve sequences with high confidence. I used maximal exact match program of BWA to map sequencing reads to reference genome followed by GATK data preprocessing (Zheng-Bradley et al., 2017, Li and Durbin, 2010, Van der Auwera et al., 2013). I also used GATK germline short variant discovery and germline copy number variant discovery pipelines to call high confidence SNPs, indels and copy number variants (Poplin et al., 2017, Van der Auwera et al., 2013). Thus, I enlighten OR gene sequence features retrieved by targeted capture sequencing in this chapter. I focus on sequencing depth of targeted capture retrieved sequences in comparison with whole genome sequencing (WGS) and whole exome sequencing (WES) data available from 1000 Genomes Project. I also presented results of short variant discovery including SNPs, insertions, and deletions, particularly I focused on novel variant discovery and presence of nonsynonymous SNPs.

2.2. Materials and Methods

2.2.1. Study Populations

I briefly introduced characteristics of my study populations in this section.

2.2.1.1. Asian Populations

I studied two of Asian populations, Japanese and Filipinos.

2.2.1.1.1. Japanese Population

The Japanese population in this study includes Honshu, Ryukyu Islanders and Hokkaido (Ainu). The Japanese population is a hybrid between the Yayoi people, farmers who migrated from the East Eurasian Continent ~3000 years ago, and the Jomon people, indigenous hunter-gatherers who were likely to be the descendants of late-Paleolithic people in the Japanese archipelago (Hanihara, 1991, Jinam et al., 2021, Adachi et al., 2021, Osada and Kawai, 2021, Koganebuchi and Oota, 2021). Ancient DNA analyses inferred that the Jomon people diverged from the current East Asians before 26000 years ago (McColl et al., 2018, Gakuhari et al., 2020). A recent genome analysis showed that local populations in Japanese archipelago exhibited varying degrees of Jomon ancestry in their genome with a geographical gradient (Watanabe et al., 2021). For example, Ainu and people in the Ryukyu islands inherit more components of the Jomon genome than people in the central part of the mainland Japan (Matsukusa et al., 2010, Koganebuchi et al., 2012b, Sato et al., 2014, Jinam et al., 2015).

2.2.1.1.2. Filipino Population

Filipino population in this study included four subpopulations of indigenous people with history of hunter-gather "Negritoes" (Aeta, Agta, Batak and Mamanwa) and three subpopulations of non-Negritoes with history of agriculturist (Tagalog, Visayan and Manobo) (Omoto, 1984, Omoto, 1981, Omoto et al., 1978). This subsistence information for each population is mentioned in 'Notes' column of Table 1. There are also reports of higher odor name proficiency in hunter-gatherer populations of East Asia, indicative of higher olfactory ability (Majid and Kruspe, 2018).

2.2.1.2. European Populations

I included four European populations in my research (Danes, Irish, Russians and Adygei). Danes make 95% population of Denmark. These people are coalesced as single kingdom since 8th century AD. Historically, they subsist on agriculture (Thurston, 1997, Osier et al., 2002). Irish people reside in Ireland since ~6000 BC and subsist on agriculture (North et al., 2000, Osier et al., 2002). Russian people make largest ethnic group to live in Russian Federation. Current study individuals are of Russian European descent and subsist on agriculture. (Halperin, 1983, Osier et al., 2002). Adygei people live in Adygei Autonomous Republic, a landlock state, of Russian Federation. They are also called as Circassians and subsist on agriculture (Wallis, 1987, Osier et al., 2002).

2.2.1.3. African Populations

Four African populations were included in this study (Biaka Pygmies, Mbuti Pygmies, Chagga and Hausa). Pygmies historically subsist through hunter-gather. Biaka Pygmies live in southern region of Central African Republic (Cavalli-Sforza, 1986, Heymer, 1980) while Mbuti Pygmies live in Democratic Republic of Congo (Putnam, 1948, Turnbull, 1983, Turnbull, 1965). Chagga people are descendent of several of various ethnic groups and live in northern Tanzania. These people subsist on agriculture and cultivate bananas, maize and coffee etc. (Moore and Puritt, 2017). Hausa is largest ethnic group of West Africa and most of these people live in northwestern Nigeria, also called as Hausaland (Greenberg, 1947). Hausa also subsists on agriculture. (Osier et al., 2002)

2.2.2. Genomic DNA Samples

Current study included 401 individuals from 18 populations of Asian-, European- and African- origins. Number of samples for each population are shown in Table 1.

The DNA samples of 54 anonymized Honshu individuals were purchased from the Japanese B Cell DNA Bank (JBCDB), as representing mainland Japanese samples. The JBCDB is managed by the Health Science Research Resources Bank (HSRRB)/ Japanese Collection of Research Bioresources (JCRB) Cell Bank in National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN). The cell sources were originated from Epstein-Barr virus transformed B-lymphoblast cell lines which were established by the Japan Biological Informatics Consortium (JBIC) collaborated with the Tokai University School of Medicine and the Institute for Genome Research of the University of Tokushima, and by the Pharma SNP consortium (PSC) collaborated with the Institute of Rheumatology of the Tokyo Women's Medical University (Kamatani et al., 2004, Ozeki et al., 2006, Sakamoto et al., 2007, Nishimoto et al., 2010, Takata et al., 2008). These samples were previously reported (Akhtar et al., 2022). Research using these samples was approved by the ethical committee at the Graduate School of Frontier Sciences of the University of Tokyo, under the approval number 20-354.

The DNA samples of 15 anonymized Ryukyu islands individuals were previously reported (Matsukusa et al., 2010, Akhtar et al., 2022) and were from blood samples collected in Miyako and Ishigaki Islands (five individuals each) and from saliva samples collected in main-land Okinawa (five individuals). Research using these samples was approved by the ethical committee at the Graduate School of Frontier Sciences of the University of Tokyo, under the approval number 20-355.

The blood DNA samples of 29 anonymized Ainu were collected by Professor Keiichi Omoto (currently Professor Emeritus) of The University of Tokyo in 1983 and 1984. The blood collection was conducted at Biratori Town, Saru District, Hokkaido, in collaboration with Professor Satakero Watanabe of Sapporo Medical University and Dr. Hiroo Kimura, the head of Biratori Eisei Kenkyujo. These samples were previously reported (Harihara et al., 1986, Harihara et al., 1988, Horai et al., 1996, Bannai et al., 1996, Bannai et al., 2000, Tajima et al., 2004, Hammer et al., 2006, Koganebuchi et al., 2012a, Nishida et al., 2012). Research using these samples was approved by the ethical committee at the Graduate School of Frontier Sciences of the University of Tokyo, under the approval number 18-278.

The anonymized blood DNA samples of the seven Filipino populations were collected by Professor Keiichi Omoto (currently Professor Emeritus) of The University of Tokyo since 1975 through 1985. The blood collection was conducted in collaboration with Dr. Joaquin S. Sumpaico and Dr. Pacifico M. Medado of Bureau of Research and laboratories, Department of Health, Manila, Philippines, Dr. Ildefonso G. Pagaran of Rural Health and Family Planning Center, Cabadbaran, Agusan del Norte, Philippines, and Dr. Francisco A. Datar of Department of Anthropology, University of the Philippines. These samples were previously reported (Omoto et al., 1978, Matsumoto et al., 1979, Omoto et al., 1981, Horai et al., 1981, Omoto, 1979, Omoto, 1980, Omoto, 1981, Omoto, 1982, Omoto, 1984, Dávila et al., 2002, Jinam et al., 2017). Research using these samples was approved by the ethical committee at the Graduate School of Frontier Sciences of the University of Tokyo, under the approval number 18-278.

The anonymized European and African samples were collected and provided by Professor Kenneth K. Kidd of Yale University School of Medicine, United States of America. The DNA samples were originated from Epstein-Barr virus transformed B-lymphoblast cell

lines and are registered in an international database ALFRED (https://alfred.med.yale.edu/alfred/index.asp) (Osier et al., 2001) and in the HGDP-CEPH Human Genome Diversity Cell Line Panel (https://cephb.fr/en/hgdp_panel.php). IDs of the samples in ALFRED are: Danes (UID PO000007H), Irish (UID PO000057M), Russians (UID PO000019K), Adygei (UID PO000017I), Biaka Pygmies (UID PO000005F), Mbuti Pygmies (UID PO000006J), Chagga (UID PO000324J) and Hausa (UID PO000097Q). Research using these samples was approved by the ethical committee at Yale University School of Medicine (HIC#1387) and that at the Graduate School of Frontier Sciences of the University of Tokyo, under the approval number 16-262.

2.2.3. OR Probe Design

The sequence sources of the OR genes for designing probes in this study are listed in Appendix II. In the appendix, the gene name is given on the basis of its chromosome number, the OR gene cluster number on the chromosome, and the gene order in the cluster (Niimura and Nei, 2003, Niimura et al., 2018). For example, *HsOR1.4.2* represents the human (*Homo sapiens*) OR gene located on the chromosome 1, in the OR cluster 4 of the chromosome, and the second gene in the cluster. Pseudogenes are indicated by suffix ‘P’ or ‘P0’ after the gene name. The gene names provided by the Human Olfactory Receptor Data Explorer (HORDE) (https://genome.weizmann.ac.il/horde/, last accessed December 27, 2021) are also given for clarity. The gene-residing chromosome or scaffold, gene position coordinates in databases, and transcriptional direction relative to the coordinates are given. Categories are also given to each gene as ‘F’ for intact (putatively functional) genes, ‘PF’ for possibly functional genes, and ‘P0’ and ‘P1’ for “nearly intact” pseudogenes as explained below. For each OR gene, 100-bp flanking sequences to the coding region (100-bp upstream and downstream regions to

gene positions in genome) were included in the probe. Repeat sequences in the flanking regions were deleted if any.

The nucleotide probes for OR genes were designed for the following sequence categories.

(1) 398 intact (putatively functional) OR genes in the human reference genome hg38 (abbreviated as hg38-intact in this study) (gene numbers 1~398 in Appendix II). These are labelled as ‘F’ in the Category column of Appendix 1. 397 genes are autosomal, and one gene is X- non-pseudoautosomal.

(2) Four sequences that are only in 'alt' (hg38-alt) (gene numbers 399~402 in Appendix II). These are labelled as ‘F (alt)’ in the Category column of Appendix II. All four are autosomal. The hg38 contains haplotype information as 'alt' that is similar to OR genes but not included as OR genes. Most of these sequences in the 'alt' are identical to or only slightly different from the OR genes in hg38 (thus are considered as alleles of the hg38-intact genes). But these four sequences are only in 'alt' and the amino acid identity of them are 89–95% to the most similar standard ones. The sequences which correspond to these four genes have been identified in HORDE (Appendix II) with 99.4%~100% identity. The possibly orthologous sequences also exist in the chimpanzee (*Pan troglodytes*) with 97.3%~98.7% identity (GenBank accession numbers XM_527318.5, XM_521959.3, XM_009440401.2, XM_001158962.3 for the gene numbers 399-402, respectively) and bonobo (*Pan paniscus*) with 97.4%~98.9% identity (XM_003829755.3, XM_003806071.2, XM_003806072.3, XM_003806807.2).

(3) 99 “nearly intact” sequences (hg38-pseudo) (gene numbers 403~501 in Appendix II). All 99 genes are autosomal. The criteria for intact gene in designing probes was that i) its open reading frame (ORF), starting with a methionine codon and ending with a stop codon, shows

a significant similarity to known OR genes and ii) there are only two or less missing amino acids in the highly conserved amino acid sites (Niimura, 2013, Niimura et al., 2018) in the alignment of Class 1 or Class 2 OR family. By extending the criteria, “nearly intact” sequences were sub-categorized as follows:

(3-1) Four sequences of “possibly intact” OR genes (Category ‘PF’ in Appendix II). These sequences are not interrupted by stop codons or frameshifts, but the number of missing amino acids in the highly conserved sites is three to five.

(3-2) 91 sequences of Category ‘P0’ in Appendix II. These sequences are interrupted by one stop codons or one frameshift, and the number of missing amino acids in the highly conserved sites is two or less (Niimura, 2013).

(3-3) Four sequences of Category ‘P1’ in Appendix II. These sequences are interrupted by one stop codons or one frameshift, and the number of missing amino acids in the highly conserved sites is three to five.

(4) 53 intact OR genes in the chimpanzee genome database Pantro3.0 with no orthologous genes in hg38 which meet (1) ~ (3) (Gene numbers 502~554 in Appendix II). These are labelled as ‘F’ in the Category column of Appendix II.

2.2.4. Neutral Reference Probe Design

85 single-copy non-protein-coding sequences (82 autosomal and three X- non-pseudoautosomal) were selected from the human reference genome hg19 assembly using the Neutral Region Explorer (Arbiza et al., 2012) as a control reference to evaluate copy number variation and selective neutrality (Appendix III). I selected regions without repeat sequences and with length longer than 1.0 kb, distance to the nearest gene longer than 0.2 centimorgan

(cM) in autosomes or 0.1 cM in the X chromosome, and minimum recombination rate 0.9 cM/Mb under the Neutral Region Explorer. At the stages of variant calling and copy number calling, these sequence positions in hg19 were converted to those in hg38 using the UCSC LiftOver (Kent et al., 2002). Because the genomic region of *hg19_NR50* was missing in hg38, the hg19 sequence were directly used for the procedures.

2.2.5. Targeted Capture & NGS

Synthesis of the in-solution biotinylated RNA baits (myBaits®) was outsourced to Biodiscovery, LLC (Ann Arbor, MI, USA) as the targeted capture probes (Gnirke et al., 2009). Each probe was 120 nucleotide (nt) length and was overlapped 60 nt with adjacent probes (i.e. 2 x tiling). An exemplary illustration of probes and an OR gene is shown in Figure 3. Construction of DNA sequencing library, targeted capture and NGS were conducted by the Centre for Health Genomics and Informatics and UCDNA Service, University of Calgary, Canada. DNA sequencing library was constructed by fragmenting genome DNA into ~500 bp and by attaching adaptor oligonucleotides to the fragmented DNA. Genomic DNA samples with distinct adaptors were pooled together in groups of six. The library was heat-denatured in the presence of adapter-specific blocking oligonucleotides. Library and blockers were dropped to the hybridization temperature, allowing blockers to hybridize to the library adaptors. Biotinylated RNA baits were introduced and allowed to hybridize to targets for several hours. Bait-target hybrids were pulled out of the solution with streptavidin-coated magnetic beads. Beads were stringently washed several times to remove non-hybridized and non-specifically hybridized molecules. These hybridization and washing conditions allow upto ~10% mismatch according to manufacturer's protocol. The captured DNA library was released from the beads and amplified by PCR using complementary adaptors. The massive

parallel sequencer used in this study was the Illumina NextSeq with 300 cycle and 150-bp paired end sequencing.

2.2.6. Bioinformatic Analysis

2.2.6.1. Processing of hg38 Sequences

The procedure is outlined in Figure 4.

(1) The PRINSEQ tool was used for quality control of raw NGS reads (Schmieder and Edwards, 2011). For the FASTQ-formatted reads with the base quality score emitted by the Illumina sequencer at each nucleotide position, any nucleotide position with quality score below 20 at either edge of a read was trimmed away. Any read with a mean quality score less than 20 and any read shorter than 20 nucleotides were also filtered out.

(2) The BWA MEM tool was used to map the quality-controlled reads to the human reference genome assembly hg38 and *hg19_NR50* as well as the 53 OR orthologous gene sequences from the chimpanzee reference genome assembly Pantro3.0. The tool is specialized to map highly similar reads on a reference genome (Zheng-Bradley et al., 2017), which is the case of OR genes because of large amount of highly similar paralogous genes.

(3) After the mapping, by applying the MarkDuplicatesSpark of the Genome Analysis Toolkit (GATK) (McKenna et al., 2010, DePristo et al., 2011, Van der Auwera et al., 2013), I identified read pairs that are likely to have originated from duplicates of the same original DNA fragments through some artefactual processes such as PCR amplification. Only a single read pair within each set of duplicates was used for the subsequent analyses.

(4) I then masked known variant sites reported in the HapMap 3.3 (Consortium, 2005, Consortium, 2007), the 1000 Genomes Omni 2.5, the high confidence SNP sites in the 1000

Genomes Phase 1 (Consortium, 2012, Consortium, 2015) and the dbsnp 138 dataset to avoid removing these sites from my variant dataset by error in the subsequent base quality score recalibration and variant discovery steps.

(5) I conducted the Base Quality Score Recalibration (BQSR) using the BaseRecalibrartor, a machine learning algorithm of GATK, to detect patterns of systematic errors in the base quality scores and correct the errors possibly originated from biochemical processes during library preparation and sequencing, from manufacturing defects in the chips, or from instrumentation defects in the sequencer.

(6) The number of reads at one nucleotide site was averaged among all nucleotide sites (sequencing depth) in an OR gene or a neutral reference region per individual using the Samtools bedcov (Li et al., 2009). The sequence depths were further averaged among carrier individuals for every gene/region or among genes/regions for every carrier.

(7) For SNPs and indels discovery, the pipeline (4.1.2.0) of GATK was used. The HaplotypeCaller was used to identify haplotypes of SNPs and indels relative to the reference sequence based on the physical linkage of variants in NGS reads per sample. Then, variant information from all samples was joined to make a variant database using the GenomicsDBImport of GATK. I then used the GenotypeGVCFs of GATK to extract multi-sample variant calls in the Variant Call Format (VCF) (Osada and Kawai, 2021).

(8) I then conducted the Variant Quality Score Recalibration (VQSR) using VariantRecalibrator and ApplyVQSR, machine learning algorithms of GATK which incorporates known variant site information. As the source information on true variants, I used the HapMap 3.3 (Consortium, 2005, Consortium, 2007), the 1000 Genomes Omni 2.5, the high confidence SNP sites in the 1000 Genomes Phase 1 (Consortium, 2012, Consortium,

2015) and dbsnp 138. Any variants within the first percentile of variant quality score log odds in the variant recalibration were filtered out.

(9) I further refined variant calls using the CalculateGenotypePosteriors of GATK which evaluates genotype likelihood as the genotype quality score for each variant sites in each study individual. Any variant sites with genotype quality score less than 20 were filtered out by the VariantFiltration of GATK. I used the VariantAnnotator of GATK to annotate possible *de novo* mutations. Upon these steps thus far, the resulted VCF dataset can be regarded highly confident and was used in downstream analyses.

2.2.6.2. Mapping & Assemblage of Non-hg38 Sequences

For 53 OR genes absent in hg38 but present in chimpanzee reference genome Pantro3.0 and for the neutral reference *hg19_NR50*, another set of mapping of all PRINSEQ-controlled reads was conducted to a custom reference set containing hg38 sequences, the 53 Pantro3.0 sequences and the *hg19_NR50* sequence. After the step (1) in the 'Processing of hg38 sequences' section, reads mapped to the 53 OR gene sequences at the step (2) were extracted using a BED file (Quinlan and Hall, 2010). Extracted reads were assembled using AbySS (Simpson et al., 2009, Jackman et al., 2017) at custom *k*-mer value of 55 which was considered best after testing several values. After assemblage, a custom BLAST was performed between 53 Pantro 3.0 OR gene sequences and their assembled sequences to confirm their orthology (Johnson et al., 2008, Coordinators, 2016). Replacing the Pantro 3.0 sequences with the assembled human sequences as references, I repeated the step (2) and went through the steps (3) to (10) except for the steps (4), (5) and (8). The steps (4), (5) and (8) are not supported for sequences which do not exist in hg38. Regarding *hg19_NR50* sequence, the steps (1) ~ (10) except for (4), (5), and (8) were conducted.

2.2.7. Statistical Analysis

All statistical analyses were performed using GraphPad QuickCalcs, Microsoft Excel for Microsoft 365 MSO, IBM SPSS Statistics 22.0 or RStudio 3.6. Mean and standard deviations were considered as measure of central tendency and dispersion on all normally distributed data. Parametric statistical tests were performed to draw inference. *p*-value less than 0.05. Parametric tests performed include one sample *t*-test, independent sample *t*-test and One way ANOVA followed by Tukey's post-hoc analysis.

2.3. Results

2.3.1. Sequence Retrieval

The 398 hg38-intact OR genes, the four hg38-alt OR genes, the 99 hg38-pseudo OR genes and the 53 Pantro3.0 OR genes were examined in 401 individuals. All of the 501 hg38 OR genes were retrieved from all or most of the 401 individuals.

On the other hand, among the 53 Pantro3.0 OR genes, 50 genes were not detected in any of the 401 individuals. One gene, *Chimpanzee_CM000325.3_5526517_5527518+*, was detected in some individuals of Ryukyu, Aeta, Batak, Mamanwa, Manobo, Chagga, Dane, Irish and Russian people in intact form. The other two genes, Chimpanzee_CM000328.3_3583176_3584120+ and Chimpanzee_CM000325.3_55413990_55414940+, were retrieved as disrupted gene. The former, Chimpanzee_CM000328.3_3583176_3584120+, was detected in some Agta, Batak, Dane, Irish, Adygei, Russian, Mbuti Pygmy and Hausa individuals. The latter, Chimpanzee_CM000325.3_55413990_55414940+, was detected in only one Mamanwa individual. Frequency of these three chimpanzee OR genes retrieval is shown in Figure 5.

Thus, total of 504 OR genes were retrieved, showing that the composition of OR genes in the standard human genome data was not necessarily representative of the OR gene composition in humans.

84 out of the 85 neutral references were retrieved from all the 401 individuals. For one region (*hg19_NR49*), NGS reads were only patchily mapped to the reference in all individuals. Thus, the *hg19_NR49* was excluded from subsequent analyses.

2.3.2. Sequencing Depth

The numbers of NGS reads at one nucleotide site averaged among all nucleotide sites (sequencing depth) in every OR gene or every neutral reference region per individual are plotted in Figure 6. Data are presented for the 500 autosomal hg38 OR genes (mean 313.9 ± 244), the 81 autosomal neutral references (mean 214.4 ± 188), the one X-chromosomal hg38 OR gene (mean 226.2 ± 185) and the three X-chromosomal neutral references (mean 117.4 ± 121) of the 401 individuals, as well as three Pantro3.0 OR genes among retrieved individuals (73.8±29). The per-site depth less than 1.0 was regarded as absence of the gene/region and were not included in the plot.

Mean depths of the autosomal OR genes and the autosomal neutral references were significantly larger than those in the 1000 Genomes Project WGS (1K WGS) (7.4) and the whole exome sequencing data (1K WES) (65.7) in which autosomal genome regions are a major component (Consortium, 2015) (Figure 7)(p-value <0.0001). One sample t-test was used for this purpose. Depths of the X-chromosomal OR gene and the X-chromosomal neutral references were lower than autosomal counterparts being consistent with chromosomal ploidy difference in male samples (Figure 7). The low depth of the three Pantro3.0 OR gene was comparable to those of X-chromosomal regions. Larger depth of OR

genes than neutral references is consistent with frequent observation of CNV in OR genes shown in Chapter 3 of this thesis.

For each of the 500 autosomal hg38 OR genes, the 81 autosomal neutral references, the one X-chromosomal hg38 OR gene and the three X-chromosomal neutral references, the sequencing depths were averaged among its carriers and are plotted in Figure 8 together with the sequencing depth of three chimpanzee OR sequences found in their carriers.

Likewise, for each carrier, the sequencing depths were averaged among the 500 autosomal hg38 OR genes, among the 81 autosomal neutral references, and among the three X-chromosomal neutral references, and are plotted in Figure 9 together with the sequencing depths of the one X-chromosomal hg38 OR gene in its carriers and that of chimpanzee OR genes in their carriers. In all plots (Figure 6, Figure 8 and Figure 9), a broad range of depths were evident. However, depths in the interquartile range and the averaged depths in the whisker range were larger or comparable to the mean depth of 1K WES (65.7) as shown in Figure 7.

Mean depth of autosomal OR genes and neutral references per individual are plotted in Figure 10. It shows that depth varies among populations possibly reflecting difference of sample quality among populations. It also shows that depth of OR genes is correlated with depth of neutral references, with OR depth being moderately higher than neutral reference depth. Mean depths of all OR genes / neutral references average across all individuals and mean depths of all individuals averaged across all OR genes / neutral references are given in Appendix IV and V respectively.

2.3.3. Variant Sites

Table 2 shows the numbers of variant sites (SNPs, insertions, and deletions relative to the references) detected in the 401 study individuals belonging to 18 populations of Asia, Africa and Europe. In the 398 hg38-intact OR genes, there were 4697 variant sites among all individuals (4504 SNPs, 50 insertions and 143 deletions) of which 480 (453 SNPs, 25 insertion and two deletions) were novel, not previously reported in HapMap, 1000 Genome or dbsnp projects. In the four hg38-alt OR genes, there were 42 variant sites among all individuals (40 SNPs and two deletions) of which all were novel. In the 99 hg38-pseudo OR genes, there were 1252 variant sites among all individuals (1187 SNPs, 20 insertions and 45 deletions) of which 261 (239 SNPs, six insertions and 16 deletions) were novel. In three Pantro3.0 retrieved OR genes, only 13 variant sites (all SNPs) were found. In the 84 neutral references, there were 4014 variant sites among all individuals (3735 SNPs, 70 insertions and 209 deletions) of which 1007 (893 SNPs, 24 insertions and 90 deletions) were novel.

Among all individuals, the number of amino-acid altering (non-synonymous) SNPs was 3316 in the 398 hg38-intact OR genes, of which 450 were novel (Table 2). That in the four hg38-alt OR genes was 30 of which all were novel.

The mean numbers of SNPs per 1000 bp (SNP density) in 398 hg38-intact OR genes (12.5 among all individuals), in four hg38-alt OR genes (11.1), in 99 hg38-pseudo OR genes (13.3), and in 84 neutral references (15.6) are also shown in Table 2. In three Pantro3.0 retrieved OR genes, SNP density per kb was 4.82 among carriers. The density was highest in neutral references, which is consistent with expectation given the difference in functional constraint on them.

Minor allele frequency of each variant site in intact OR genes, OR genes only in Alt and 99 OR pseudogenes was also calculated. 3615 variant sites from intact OR genes and 32

variant sites from OR genes only in Alt were found to be in minor allele frequency less than 5%. 992 variant sites out of 1252 in OR pseudogenes were also found to present in minor allele frequency of less than 5%. 3198 variant sites in neutral references were observed in minor allele frequency of less than 5%. This lower allele frequency argues population specific variations among OR repertoire and merit of my approach to retrieve many minor alleles. Frequencies of these minor allele frequencies in each OR gene category and neutral references is shown in Table 2.

2.4. Discussion

In this study, I applied a targeted capture approach paired with short-read NGS to examine genetic variation in 401 individuals, belonging to 18 ethnic groups of Asian-, European- and African-origins, for 398 intact OR genes, four intact but unannotated OR genes and 99 "nearly-intact" OR genes in the human reference genome hg38 and 53 intact OR genes in the chimpanzee genome database Pantro3.0 with no orthologous genes in hg38. I found that the targeted capture by probes with diversity-oriented design was far more effective than the whole-genome approach to retrieve OR genes and achieve high-depth sequencing and thus to reveal intact/disrupted polymorphisms and CNVs for the OR multigene family. I also found that the composition of OR genes in the reference human genome hg38 was not necessarily representative of those in humans and that many of the OR gene variants identified in this study were not previously reported, implying the potential for higher perceptual variation in humans than ever thought.

I retrieved three Pantro3.0 OR genes (*Chimpanzee_CM000325.3_5526517_5527518+*, *Chimpanzee_CM000328.3_3583176_3584120+ and*

Chimpanzee_CM000325.3_55413990_55414940+) from different individuals as intact and pseudogenes. Although rare, their existence demonstrates the importance and effectiveness of population surveys using targeted capture based on a probe set that includes genes from close evolutionary relatives.

I noted that the sequencing depths of the autosomal OR genes and the autosomal neutral references were roughly in line with male/female ratio of those of the non-pseudoautosomal X-chromosomal OR gene and the neutral references, respectively (Figure 7), which is consistent with their ploidy difference among genders.

I also noted that the sequencing depths in this study [the mean depths 73.8 (OR-P) ~ 313.9 (OR-A) (Figure 7)] were significantly larger than that of the WGS (7.4) from the 1000 Genome Project. Because many multigene families assemble poorly with low coverage in WGS databases using short-read NGS (Sims et al., 2014, Yohe et al., 2020) and because OR genes represent only ~0.0125 % of the whole genome in mammals (Lane et al., 2001, Mainland et al., 2013), the successful high coverage by targeted capture in this study should improve reliability and efficiency of identification of OR multigene family and its genetic variation. I also identified a wealth of novel variant sites in my dataset including nonsynonymous SNPs which were not previously reported in HapMap, 1000 Genome or dbsnp projects.

SNP density calculated by these variant sites as shown in Table 2 also served as a measure to inspect bioinformatic processing of sequencing data. Highest SNP density was observed in neutral references (15.6) followed by OR pseudogenes (13.3) and then intact OR genes (12.5). Neutral references are non-protein-coding single-copy parts in human genome. They are expected to be under less functional constraint and under less negative natural selection to mutations. In contrast, intact OR genes are expected to be under functional

constraint, and non-synonymous mutations could be disruptive and under negative natural selection. Thus, under the same number of sequences, SNP density in intact OR genes is expected to be lower than that in neutral references. Disrupted OR genes were intact in some evolutionary past. Thus, SNP density of disrupted OR genes is expected to lie between intact OR genes and neutral reference. Results shown in Table 2 are consistent with this expectation. Thus, the bioinformatic processing of NGS reads in this study can be regarded appropriate.

Based on the high depth OR gene sequences retrieved in my study and many novel variant sites observed, I conclude this chapter with merits of targeted capture sequencing over whole genome sequencing.

CHAPTER 3

Segregating Pseudogene Polymorphism and Copy Number Variation of Olfactory Receptor Genes in Diverse Human populations

3.1. Introduction

Humans are generally considered visually oriented animals and olfaction may not be considered of as importance as vision. This view is not only linked to humans but also to other primates especially haplorhines (Niimura et al., 2018). Haplorhines are simple nose primates which consists of catarrhines [hominoids (humans and apes) and cercopithecids (African/Asian monkeys)], platyrrhines (American monkeys) and tarsiers in contrast to strepsirrhines with curved noses which consist of lemurs and lorises. Size of OR gene repertoire among strepsirrhines is nearly twice of haplorhines. This inter-taxonomic difference of OR gene repertoire is considered to be shaped by difference of nasal anatomy, habitat and phylogeny (Niimura et al., 2018).

Humans are diverse in living environment, historical subsistence, and culture, which could also have influenced the OR gene repertoire. Current OR gene repertoires have been shaped by evolutionary processes with natural selection and random genetic drift. Thus, studying OR repertoire across populations with diverse living environments, historical subsistence, and culture such as current study populations as described in section 2.2.1 in Chapter 2 of this thesis could be helpful to understand how OR repertoire has been shaped. The first question is whether and how OR gene repertoire is variable among populations with different living environments, historical subsistence, and culture.

OR gene family is known to exhibit segregating pseudogene polymorphism and CNV (Niimura et al., 2018, Olender et al., 2013). A segregating pseudogene is the gene that exhibits both intact and disrupted alleles while CNV polymorphism of a gene locus exhibits deletions of the locus and/or duplications of the locus as alleles (Olender et al., 2013). Together with SNPs, these genetic variations could contribute significantly to olfactory sensory variation in humans.

In this chapter, I investigated segregating pseudogene polymorphism (intact/disrupted allele polymorphism) and CNV of the OR gene family among individuals within a population and between populations.

3.2. Materials and Methods

3.2.1. Study Individuals

All 401 individuals belonging to 18 populations as described in Table 1 were included in this analysis.

3.2.2. Sequencing Data

Sequences retrieved by targeted capture sequencing as described in 2.3.1 were included. This data includes 501 hg38 OR genes and 3 chimpanzee OR genes retrieved in study individuals.

3.2.3. Intact/Disrupted Gene Allele Composition

Intact/disrupted allele composition of all 401 study individuals were determined. A consensus sequence of each gene was constructed using bcftools consensus (Li, 2011, Danecek et al., 2021). Intact OR gene sequences are ~930 bp long (encoding ~310 amino acids long) (Niimura, 2013). EMBOSS getorf was used for finding ORF (Rice et al., 2000). I regarded any gene with maximum ORF shorter than 750 bp (encoding shorter than 250 amino acids long) as a disrupted gene unlikely to form a functional structure by following Niimura's criteria (Niimura, 2013). I also examined if an OR sequence is capable of forming 7-transmembrane (7-TM) structure, a hallmark of GPCRs by using TMHMM 2.0 (Krogh et al., 2001). I regarded any gene with an ORF of 750 bp or longer but without proper 7-TM

structure as a disrupted gene. Thus, only OR genes with an ORF of at least 750 bp and capable of forming a 7-TM structure were considered as intact genes. For intact OR genes, non-synonymous variants were identified using SnpEff (Cingolani et al., 2012).

3.2.4. CNV Analysis

The germline copy number variant discovery pipeline (4.1.9.0) of GATK was used to identify CNV of the hg38-OR genes and of hg38-liftover neutral references among individuals using the Bayesian model (Van der Auwera et al., 2013). In the first step, I instructed the PreprocessIntervals of GATK which part of the hg38 data was my study targets. I used the CollectReadCounts of GATK to count depth of each target region for each individual for CNV analysis. In the next step, I used the AnnotateIntervals and the FilterIntervals of GATK to remove genes/regions judged as outliers in terms of high GC content because higher GC content can facilitate targeted capture and cause higher depth regardless of its copy number.

I used the DetermineGermlineContigPloidy of GATK to pre-set ploidy level for each chromosome based on chromosomal locality and depth information of our study gene/region using the default prior ploidy probabilities provided by GATK for autosomes and X chromosomes. I used the GermlineCNVCaller of GATK to estimate copy number state (ploidy) of each OR gene and neutral reference in terms of 0*n*, 1*n*, 2*n*, 3*n*, and so on, for every individual based on its sequencing depth relative to the mean depth among all OR genes and neutral references from the same individual. PostprocessGermlineCNVCalls was then used to extract copy number state of each sequence in VCF file format. The GATK pipeline treats the mean depth as the single-copy diploid standard because majority of the genes and genome regions are autosomal. When the depth of a gene/region of interest is evaluated not to significantly differ from the mean depth, its copy number is estimated to be

2*n*. The GATK pipeline for the CNV analysis was not applied to the non-hg38 sequences for simplicity.

3.3. Results

3.3.1. Disruption of OR Genes

Among the 504 OR genes retrieved (section 2.3.1 in Chapter 2) in this study, 134 genes were polymorphic with disrupted and intact alleles (segregating pseudogenes) in the 401 individuals. Among 134, 128 genes were the hg38-intact OR genes and were segregating pseudogenes in all populations. The other 270 hg38-intact OR genes were intact in all individuals. One of 134 was a hg38-alt OR gene (Human_chr11_JH159136v1_alt_193979_194908+) was segregating pseudogene only in Aeta and Agta populations of Philippines. Three of 134 were the category P1 (a "nearly intact" pseudogene category) (*HsOR7.6.17P0*, *HsOR14.1.9P0* and *HsOR14.1.28P0*) and were segregating pseudogenes in all populations. One of 134 was also the category P0 (*HsOR1.5.11P0*) and was segregating pseudogene only in one Hausa study individual. One of 134 was the category PF ("possibly functional") (*HsOR19.3.8P*) and was segregating pseudogenes in all populations. The other 94 hg38-pseudo OR genes were disrupted in all individuals.

Frequencies of disrupted alleles of these 134 OR segregating pseudogenes are shown in Figure 11 among of all samples. While frequencies of disrupted alleles among all samples were relatively high (33~70%) in the four of the five genes from the hg38-pseudo category (*HsOR7.6.17P0*, *HsOR14.1.9P0*, *HsOR14.1.28P0* and *HsOR19.3.8P*), those of many hg38-intact OR genes were also high (over 10% in 14 genes and maximally 64.5%) (Figure 11). On

the other hand, 109 out of 134 polymorphic segregating OR genes showed disrupted alleles at 5% or less. Most of these rare disrupted alleles were observed in only one population.

Disrupted allele frequencies were subdivided into 18 study populations. Many of the OR genes for which disrupted allele frequencies are low in Figure 11 are found to exhibit intact/disrupted allele polymorphism only in one population or group. Disrupted allele frequencies of OR genes disrupted only in Ainu, Honshu and Ryukyu populations of Japan are shown in Figure 12. Similar disrupted allele frequency incidence is shown is Negrito, non-negrito Filipino, European and African study populations in Figure 13, Figure 14, Figure 15 and Figure 16 respectively.

The number of OR genes which were homozygous with intact alleles was counted for every individual. The distribution of individuals on this number in each population is shown in Figure 17. One way ANOVA was conducted to draw statistical inference followed by Tukey's post-hoc analysis. ANOVA revealed significant difference among mean number of homozygous intact OR genes are highest in Asian populations followed by European populations and then African populations ($p < 0.0001$).

3.3.2. Copy Number Variation

Three OR genes (*HsOR8.1.1*, *Human_chr6_GL000252v2_alt_682794_683756+* and *Human_chr11_JH159137v1_alt_18234_19151+*) and one neutral reference (*hg19_NR77*) were excluded from CNV evaluation due to their high GC content (details in methods above, section 3.2.4). Regarding the 80 autosomal neutral references, all individuals were estimated to be 2*n* as expected for the single-copy standard, while three X-chromosomal neutral references were also estimated to be 1*n* in all males and 2*n* in all females as expected except one male Honshu (X chromosomal sequences: 2n) individual. As Honshu samples were

obtained from cell lines thus this deviation was interpreted as chromosomal aberration arose during cell culture.

Regarding the autosomal OR genes, the copy numbers varied among individuals in 176 genes. The genotype 0*n*, the gene-deletion homozygote, was observed in 37 OR genes. The genotype 1*n*, the gene-deletion/normal heterozygote, was observed in 47 OR genes. The genotype 3*n* was assumed to be gene-duplication/normal heterozygote but not more complicated gene-triplication/gene-deletion combination and was observed in 114 OR genes. The genotype 4*n* was likewise assumed to be gene-duplication homozygote but not more complex gene-triplication/normal heterozygote or gene-quadruplicate/gene-deletion heterozygote and was observed in 14 OR genes. The genotype 5*n* was assumed to be gene-duplication/gene-triplication heterozygote but not more complex gene-quadruplication/normal heterozygotes or pentaplication/deletion heterozygote and was observed in eight OR genes. The genotype 2*n* was simply assumed to be normal homozygote but not more complex gene-duplication/gene-deletion heterozygote.

The CNV allele frequencies of autosomal OR genes are shown in Figure 18. Allele frequencies are shown in chromosomal and gene cluster orders. Gene cluster names are shown beside gene names when two or more than two genes were found to exhibit CNV in one cluster. Among all samples, gene-deletion allele was observed in 70 OR genes, gene-duplication allele was observed in 120 OR genes and gene-triplication allele was observed in eight OR genes. It was noted that the eight OR genes with a triplication allele also carried deletion and duplication alleles (Figure 18). It was also noted that OR genes with CNV were often found in the same cluster of OR genes thus closely located in the genome (Figure 18). 164 out of 176 OR genes showed CNV allele frequency less than 5% and most of these rare CNVs were observed either in a single population or in close geographical regions.

CNV allele frequencies were further broken down to 18 study populations. These results confirmed population specificity of some CNV alleles or specificity to a particular study group. One example of this is two CNV exhibiting OR genes present only in hg38 Alt sequences. These genes showed CNVs only in Asian populations but not in European or African populations. CNV allele frequencies of CNV exhibiting OR genes in each population are given in Figure 19, Figure 20, Figure 21, Figure 22 and Figure 23.

3.4. Discussion

In this study, I showed that 128 of 134 segregating pseudogenes were in the hg38-intact OR gene category and one was in hg38-alt category. In HORDE, there are 55 segregating pseudogenes among the 501 hg38 OR genes included in this study. Forty three of the 55 segregating pseudogenes are in the hg38-intact category, and one is in hg38-alt category. The frequency of disrupted alleles in many hg38-intact OR genes in this study were high (over 10% in 11 genes) (Figure 11). Thus, majority of segregating pseudogenes are in the intact OR gene categories in hg38 (hg38-intact and hg38-alt).

On the other hand, a small fraction (five genes) of the 134 segregating pseudogenes was in the hg38-pseudo OR gene category (*HsOR1.5.11P0, HsOR7.6.17P0, HsOR14.1.9P0, HsOR14.1.28P0,* and *HsOR19.3.8P*). Among these genes, intact alleles are not minor. Literature survey revealed an hg38-pseudo OR gene, *HsOR14.1.28P0,* a segregating pseudogene in my research, is reported to be expressed in the olfactory mucosa in Africans and Europeans by an autopsy study (Verbeurgt et al., 2014). *HsOR14.1.28P0* has also been reported to bind isovaleric acid, a carboxylic acid present in apple juice, cheese and soy milk (Verbeurgt et al., 2014).

Ligands for 90 ORs are listed in Appendix I by my literature search. 89 OR of those are included in this study. Out of these 89 ORs, 25 are found to be segregating pseudogenes of which disruptive allele frequency is low (0.1% – 4.1%) except *HsOR14.1.28P0.* *HsOR14.1.28P0,* a pseudogene in human reference genome, was found to have an intact allele frequency of 30.4%. This gene is also reported to bind isovaleric acid (Verbeurgt et al., 2014). These observations support the view that OR gene repertoire is highly variable, and the human reference genome database represents only one example of various OR gene repertoire in human populations.

Among the 134 segregating pseudogenes in this study, 112 genes are not reported as segregating pseudogenes in HORDE. On the other hand, among the 55 segregating pseudogenes in HORDE within the 501 hg38 OR genes, 33 genes are not segregating pseudogenes in my study: 24 of 33 genes are intact in all individuals and 9 genes are disrupted in all individuals. This observation also supports the notion that intact/disrupted composition of human OR genes is highly variable.

Among the 134 segregating pseudogenes in this study, 22 genes are reported as segregating pseudogenes in HORDE as well. Thus, intact/disrupted polymorphism of these genes would be worldwide. *HsOR11.3.63* is one of such genes. It was found to exhibit intact/pseudogene polymorphism in almost all study populations as shown in Figure 12, Figure 13, Figure 14, Figure 15, and Figure 16. The frequency of disrupted allele is high (64.5%) in my samples (Figure 11). This gene is reported to be disrupted in Denisovans and Neanderthals (Hughes et al., 2014) and is intact in the chimpanzee genome database Pantro3.0. Thus, the pseudogenization of this gene appears to be ancient, at the time of hominin common ancestor. Worldwide distribution of its intact and disrupted alleles implies a sensory role of the presence/absence variation of this OR as well as of the other 21 genes.

While segregating pseudogenes are distributed throughout chromosomal OR gene clusters, I noted that some of the segregating pseudogenes appeared to be closely located in the same clusters (Figure 11). However, disruptions occur mainly due to small indels and non-sense nucleotide changes, a possible cause of systematic incidence of these mutations is currently unclear.

I found CNV in 176 autosomal OR genes. These CNVs were created by gene-deletion, gene-duplication, and gene-triplication alleles. In most cases, CNV was due to either gene-deletion allele or gene-duplication allele. However, genes with gene-triplication allele were also found to carry gene-deletion and gene-duplication alleles (Figure 18). I also noted that genes with CNV were often found in the same OR gene clusters (Figure 18). CNVs were observed in both hg38-intact and hg38-pseudo OR gene categories. Frequencies of CNVs appear not to be correlated with these categories. These observations could suggest a chromatin-structural cause to facilitate CNV. Frequencies of CNVs were variable among OR genes. Thus, CNV could also contribute to olfactory variation.

The targeted capture by probes with diversity-oriented design is far more effective than the whole-genome approach to retrieve OR genes and achieve high-depth sequencing. Because of the large variation and the novelty of the variation of the OR gene family found in this study, olfactory perceptual variation in humans could also be larger than previously imagined, supporting a notion that everyone experiences their own unique "flavor world" (McRae et al., 2013).

In vitro expression assays (Zhuang et al., 2009, Adipietro et al., 2012) of the OR genes featured above for possible odorant ligands could promote our understanding of olfactory perceptual variation. The novel nonsynonymous SNPs should also be an important source for functional tests that seek to explore olfactory perceptual variation within and

between populations. Population-specific segregating pseudogenes and CNV are also indicative of population specific olfactory sensation. These results inspire further investigation on OR repertoire using population genetics tools to unveil underlying adaptive evolution.

CHAPTER 4

Population Differentiation of Olfactory Receptor Gene Repertoire in Humans

4.1. Introduction

In the previous chapters, I revealed how much OR gene repertoire is variable among populations in terms of non-synonymous SNPs, segregating pseudogene composition and CNV. These variants exist in both hg38 intact and pseudogene categories. Incidence of variants varies from gene to gene. Allele frequency composition of some OR genes are variable among populations while that of other OR genes are rather common among populations.

Incidence of non-synonymous SNPs is suggestive of perceptual variation of odors. Intact/disrupted and copy number variation polymorphisms may have similar functional implications. At first, I found many SNPs with minor alleles and then I found many OR genes exhibit intact/pseudogene or CNV polymorphism but not as minor alleles. This phenomenon raised a question that how OR repertoire is variable across populations and how these variation patterns are present in different populations. In Chapter 3, I found number of homozygous intact OR genes are higher among Asian populations followed by European and African populations respectively. In this chapter, I utilised SNP and sequence data using precise genetic approaches to look deeper into it.

In this chapter, I utilized population genetics tools to unveil population differentiation of OR gene compositions more in detail. I examined population structure using OR sequences and compared it with neutral references. I also investigated which OR genes are more or less differentiated among populations and discussed natural selection behind the variation.

4.2. Materials and Methods

4.2.1. Materials

Variant calls obtained for hg38-ORs (501), and neutral references (84) and their consensus sequences were used for analyses. None of the individual was excluded from this analysis.

4.2.2. Population Differentiation

Two analyses were performed for population differentiation.

4.2.2.1. Principal Component Analysis

Principal component analysis (PCA) was performed for OR genes and neutral references to visualize their genetic variation among all individuals. This analysis was conducted in RStudio using Adegent package (Jombart, 2008). Variant data of study individuals contained in a VCF file as obtained at step (9) of 2.2.6.1 was used for this analysis. vcfR package was used to import VCF file in RStudio (Knaus and Grünwald, 2017).

4.2.2.2. Inter-Population Genetic Distance

The pairwise inter-population genetic distances were calculated using d_{xy}, d_A, and pairwise N_{ST} measures averaged over all hg38-intact OR genes and averaged over all neutral references (Nei and Li, 1979, Nei, 1982, Nei and Kumar, 2000). Phylogenetic trees were constructed using the neighbor-joining method based on these distance measures (Saitou and Nei, 1987).

- d_{xy}

d_{xy} is the average pairwise nucleotide difference between a sequence from a population (x) and a sequence from another population (y). d_{xy} was calculated for each hg38 intact OR gene and each neutral reference for two given populations using pixy (Korunes and Samuk, 2021). Then, d_{xy} values were averaged over all hg38 intact OR genes and over all neutral references for the two given populations.

- ***d_A***

d_A is given by the formula:

$$d_A = d_{xy} - \frac{d_x + d_y}{2}$$

d_x (π_x) and d_y (π_y) are the average nucleotide differences between two sequences in a population (x) and in another population (y), respectively, i.e., nucleotide diversity of population x and population y, respectively. d_A was calculated for each hg38 intact OR gene and each neutral reference for two given populations using pixy (Korunes and Samuk, 2021). Then, d_A values were averaged over all hg38 intact OR genes and over all neutral references for the two given populations.

- **Pairwise *N_{ST}***

Pairwise N_{ST} is given by the formula:

$$\text{Pairwise } N_{ST} = \left(d_{x+y} - \frac{dx+dy}{2}\right) \Big/ d_{x+y}$$

d_{x+y} is the average nucleotide differences between two sequences in one fused population consisting of members of populations x and y. Pairwise N_{ST} was calculated for each hg38

intact OR gene and each neutral reference for two given populations in RStudio using R package Statistical Analysis of Mixed Ploidy Populations (Pembleton et al., 2013). Then, Pairwise N_{ST} values were averaged over all hg38 intact OR genes and over all neutral references for the two given populations.

4.2.3. N_{ST} Analysis for Population Differentiation

N_{ST} is a population differentiation coefficient equivalent to F_{ST} or G_{ST} for nucleotide sequence data (Nei and Kumar, 2000) and is given by the formula:

$$N_{ST} = \frac{\pi_T - \bar{\pi}_S}{\pi_T}$$

π_T is the average nucleotide differences between two sequences in one fused (total) population consisting of members of all populations (subpopulations). π_S is the average nucleotide differences in a subpopulation. $\bar{\pi}_S$ is averaged π_S among all subpopulations. N_{ST} was calculated for each hg38 intact OR gene and each neutral reference using a custom script.

Below are the steps for calculation of π_T, $\bar{\pi}_S$ and N_{ST} using Linux *shell*, GATK SelectVariants and VCFTools:

- **π_T Calculation**

A. Linux *Split* command was used to split a BED file line by line, thus a separate BED file for each OR gene or neutral reference sequence.

B. Linux *expr* command was used to define target region length for N_{ST} calculation.

C. GATK's SelectVariants tool was used to extract a separate VCF file for each gene or neutral reference sequence containing 401 individuals from all study populations.

D. VCFTools's 'Window Pi' function was used to calculate nucleotide diversity (π) in each target region at sliding step of 1 base for length calculated at step B. Sliding steps were chosen because VCFTools is designed for whole genome data and usually analysis starts at 1st base pair, but this is not the case for targeted regions. So, sliding steps were designed to analyze multiple regions possible regions which will also include targeted gene or neutral reference region.

E. Linux *grep* command was used to extract π of each targeted region against given BED file. This was π_T.

- **$\bar{\pi}_S$ Calculation**

A. Steps A through E were repeated for all 18 study populations.

B. Results of 18 study populations were sorted using Linux *sort* command.

C. Mean of π calculated for each population is calculated. This was $\bar{\pi}_S$.

- **N_{ST} Calculation**

N_{ST} was calculated for each OR gene and neutral reference by inputting π_T and π_S using formula above.

4.2.4. Tajima's *D* Analysis

Tajima's *D* evaluates the difference of nucleotide diversity and the number of SNPs adjusted by the number of sequences for each gene or genome region (Tajima, 1989).

Tajima' $D = \frac{d}{\sqrt{\hat{V}(d)}}$ $\quad d = \pi - s \Big/ \sum_{i=1}^{n-1} \frac{1}{i}$ $\quad s$ is the number of SNPs per site.

To evaluate if balancing selection is operating on OR genes to maintain allelic diversity in a population, I examined if Tajima's *D* values of OR genes are distributed to larger value than those of neutral references. I calculated Tajima's *D* for each OR gene and neutral reference sequence using a custom shell script utilising GATK and VCF-kit utilities (Van der Auwera et al., 2013, Cook and Andersen, 2017). Custom scripts worked in following steps given below:

A. Linux *Split* command was used to split a BED file line by line, thus a separate BED file for each OR gene or neutral reference sequence.

B. Linux *expr* command was used to define target region length for Tajima's *D* calculation.

C. GATK's SelectVariants tool was used to extract a separate VCF file for each gene or neutral reference sequence.

D. VCF-kit's Tajima function was used to calculate Tajima's *D* in each target region at sliding steps of 50 bases for length calculated at step B. Sliding steps were chosen because VCF-kit is designed for whole genome data and usually analysis starts at 1st base pair but this is not the case for targeted regions. So sliding steps were designed to analyze multiple possible regions which will also include targeted gene or neutral reference region.

E. Linux *awk, grep, sort* and *head* commands were used to curate Tajima's *D* of targeted region for each OR or neutral reference sequence.

F. All curated Tajima's *D* values were combined using Linux *cat* command and interpreted using appropriate statistical tests as described below.

I confirmed by the quantile-quantile (QQ) plot that the distribution of Tajima's *D* values was not largely deviated from normal distribution (Figure 24). Then, I performed a parametric independent sample *t*-test to compare the mean Tajima's *D* values between OR genes and

neutral references. I also performed a parametric one-sample *t*-test to compare the mean Tajima's *D* values with 0.

4.3. Results

4.3.1. SNP-Based Inter-Individual Diversity of OR Genes and Neutral References by PCA

PCA was performed to differentiate study individuals based on SNP repertoires of OR genes and neutral references. 398 hg38 intact OR genes and 84 neutral reference sequences were used for this analysis. Population structure of neutral references can be regarded as a baseline indicator of population differentiation which have been driven by random genetic drift and demographic effects (such as population size change, migration, admixture, population split etc.). Because random genetic drift and demographic effects shape the genetic variation of OR genes as well, basic diversity pattern by PCA of OR genes is expected to be similar with that of neutral references but would show deviations from it if natural selection on OR genes operates.

Distribution of individuals by PCA is shown in Figure 25. Figure 25A is for 398 hg38 intact OR genes while Figure 25B is for neutral references. Individuals were more spread in each subpopulation and were intermingled between subpopulations in neutral references while individuals were more aggregated by subpopulation in OR genes. In other words, individuals of the same subpopulation tend to have a similar pattern of genetic variation in OR genes. More spread of individuals and less differentiation in ORs may also be due to lower number of SNPs in neutral references than ORs. However, to counter such biasness, I performed other genetic measures.

I examined inter-population genetic differentiation measure, pairwise N_{ST}, and genetic distance (nucleotide differences) measure d_{xy} and d_A for hg38-intact OR genes and references (Table 3, Table 4 and Table 5). Phylogenetic tree reconstructed using pairwise N_{ST} genetic differentiation measure is shown in Figure 26. The size of the tree of OR genes is much smaller than that of neutral references. Phylogenetic trees were reconstructed using d_{xy} and d_A by neighbor-joining method (Saitou and Nei, 1987) (Figure 27 and Figure 28).

4.3.2. N_{ST} Population Differentiation Analysis

In order to compare the overall level of population differentiation between OR genes and neutral references and to identify OR genes which are the most and the least population differentiated, I conducted N_{ST} population differentiation analysis. N_{ST} values of some OR genes were negative. Such negative N_{ST} values were treated as zero as they were not an indication of genetic differentiation.

Distribution of N_{ST} values of hg38-intact ORs and neutral references are shown in Figure 29 as boxplots. Overall, N_{ST} is significantly smaller in ORs than in neutral references (p-value = 0.0109). This smaller N_{ST} in OR genes in comparison with neutral reference indicates smaller differentiation of OR genes than of neutral references. This is also in line with PCA results where individuals were more similar in OR genes than in neutral references in each population in terms of SNP pattern. This is also consistent with the smaller population tree of OR genes than of neutral references.

4.3.3. Tajima's D Analysis

I examined Tajima's *D* of hg38 intact and pseudogenes. Then I also examined Tajima's *D* of neutral references. Tajima's *D* of OR intact and pseudogenes and neutral

references is shown in Figure 30. It is observed that mean Tajima's *D* was higher for intact OR genes followed by OR pseudogenes and then neutral references. I applied independent sample *t*-test to compare if mean Tajima's *D* of intact OR genes is statistically higher than OR pseudogenes or neutral references. Independent sample *t*-test revealed Tajima's *D* of intact OR genes is significantly higher than neutral references (*p*-value=0.0003). This significantly higher Tajima's *D* is then interpreted as balancing selection signature in intact OR genes. This balancing selection signature is then interpreted as maintenance of heterozygosity by OR genes in relation to environment.

I examined previous results with Tajima's *D* for validation purposes. So I found Tajima's *D*'s balancing selection signature is consistent with previous results where pairwise N_{ST} revealed a smaller phylogenetic tree than neutral references indicative of lower differentiation within OR genes due to maintenance of heterozygosity. Similar results were obtained by N_{ST} population differentiation analysis where OR genes were less differentiated than neutral references.

Among intact OR genes, *HsOR11.3.63, HsOR11.3.84, HsOR11.18.33, HsOR9.4.8* and *HsOR3.3.6* were top outlier in Tajima's *D* distribution as showed in Figure 30. *HsOR15.1.10P0* and *HsOR11.18.22P* showed highest Tajima's *D* in OR pseudogene category. Three neutral references, hg19_NR12, hg19_NR23 and hg19_NR31, outliers in Tajima's *D* distribution of neutral references. These results were further broken down to each study population and are represented in Figure 31, Figure 32 and Figure 33.

4.4. Discussion

I examined genetic population structure of OR genes as compared with neutral references using principal component analysis (PCA). PCA is a dimension reduction method commonly

used in statistics while handling big data. It extracts patterns of data and then can be useful to determine important aspects. I applied PCA on my SNP data of 18 study populations. In Asian, European, and African populations, individuals were less scattered in ORs than in neutral references. On the other hand, in neutral references, between Asian and European populations, individuals were more scattered and the two were intermingled. This intermingling pattern seems to be due to lower number of SNPs in neutral references as compared to OR genes as PCA is a statistical method based on SNPs. African populations were distinct from Asian and European populations in both ORs and neutral references, being consistent with the "Out-of-Africa" population migration history. Within African populations, in ORs, individuals were less scattered than neutral reference in each subpopulation, being consistent with the pattern in Asian and European populations. This observation is indicative that individuals tend to have similar SNP patters in ORs than in neutral references.

The genetic distance analysis between populations showed that hunter-gatherer "Negritos" (Aeta, Agta and Batak) and agriculturist populations of Philippines were separated in OR genes while they were more mixed in neutral references. Majid and Kruspe (2018) found significantly higher odor name proficiency among east Asian hunter-gatherer populations in contrast to neighboring agriculturist populations (Majid and Kruspe, 2018). This could also be consistent with maintaining a similar set of OR gene variation among hunter-gathers.

Among Negrito, Mamanwa was not clustered with Aeta, Agta and Batak. Among both neutral references and ORs, Mamanwa position was distinct from rest of negrito populations which is in line with whole genomes SNP analysis (Jinam et al., 2017) where relatively deep divergence of Mamanwa from other Negritos are reported. Role of historical

subsistence in shaping olfactory receptor repertoire tells us importance of olfaction in food gathering strategies. Several observations about relationship between olfaction and foraging behaviors already have been noticed among primates and other mammals (Hiramatsu et al., 2009, Zhang et al., 2014).

Regarding the four OR genes highly differentiated among study populations, this high differentiation is considered as adaptation to their local environments. This adaptation may have been linked to sense specific odors and/or linked to odor preference or likeliness etc. Ligands for these four OR genes are not known yet so making interpretation a bit difficult at this time. Ligands for 90 known ORs are given in Appendix III. The *in vitro* functional assays in future to unveil associated ligands will help to determine underlying process.

On the other hand, many OR genes were also found to be not differentiated among study populations. 153 OR genes had zero N_{ST}. This zero N_{ST} is suggestive of balancing selection. To cross-check this balancing selection indication, Tajima's D of OR genes is examined in line with N_{ST}. OR genes with highest Tajima's D had very low N_{ST} (e.g., zero), an endorsement to balancing selection indication.

I picked up some interesting OR genes and tabulated them in Table 6 as featured genes. I added gene expression, ligand, disrupted allele frequency, CNV allele frequency, N_{ST} and Tajima's *D* information for each featured OR gene. Ligands for most of these featured OR genes are unknown yet and those genes show intact/disrupt polymorphism, CNV polymorphism, high differentiation across populations and/or balancing selection signatures. These genes could be a choice for future *in vitro* functional assays.

In conclusion, genetic variation of many OR genes appears to be maintained by balancing selection among human populations. This was exemplified in the case of Negritoes and non-negritoes where historical subsistence could have shaped OR repertoire. Beside this

pattern, there are a few OR genes which appears to be specific to a local population. Future *in vitro* functional assays for ligands and investigation of expression pattern could reveal functional validation for the pattern of genetic variation revealed in this study. Further, extension of Tajima's *D* analysis to other populations will also help to understand balancing selection operation.

CONCLUSIONS

Targeted capture followed by massive parallel sequencing is a better approach to study multigene families. This approach revealed high depth OR sequences in my research. Targeted capture sequencing is also a better choice in terms of efficient sequencing data volume management. OR sequences revealed that intact/disrupted allele composition is highly variable among study individuals and populations. Many novel variants including SNPs (especially non-synonymous SNPs), insertions, deletions, and segregating pseudogenes have been retrieved in this thesis. I also discovered deletion, duplication, and triplication alleles as CNV among intact and pseudo OR genes. I found that many of human OR genes have been possibly under balancing selection in a way that genetic variation pattern is maintained. I also noted that genetic variation of a few OR genes is population-specific possibly adapting to their local environments.

TABLES

Table 1: Number of individuals of each population

Country/Region	Population	N	Subsistence	Notes
Asia	**Ainu**	29	Hunter-gatherers	
	Honshu	54	Agriculturists	
	Ryukyu	15	Hunter-gatherers	
	Tagalog	22	Agriculturists	Non-Negrito Filipino
	Visayan	18	Agriculturists	Non-Negrito Filipino
	Manobo	28	Agriculturists	Non-Negrito Filipino
	Aeta	56	Hunter-gatherers	Negrito
	Agta	17	Hunter-gatherers	Negrito
	Batak	19	Hunter-gatherers	Negrito
	Mamanwa	23	Hunter-gatherers	Negrito
Europe	**Dane**	15	Agriculturists	
	Irish	15	Agriculturists	
	Adygei	15	Agriculturists	
	Russian	15	Agriculturists	
Africa	**Chagga**	15	Agriculturists	
	Hausa	15	Agriculturists	
	Biaka Pygmy	15	Hunter-gatherers	
	Mbuti Pygmy	15	Hunter-gatherers	
Total		401		

Table 2: Number of variant sites and SNP density among study individuals

	398 hg38-intact OR genes	4 hg38-alt OR genes	99 hg38-pseudo OR genes	3 Pantro3.0 retrieved OR genes	84 neutral references
SNP	4504 **(453)**	40 **(40)**	1187 **(239)**	13 **(13)**	3735 **(893)**
Non-synonymous SNP	3316 **(450)**	30 **(30)**	NA	-	NA
Insertions	50 **(25)**	0	20 **(6)**	-	70 **(24)**
Deletions	143 **(2)**	2 **(2)**	45 **(16)**	-	209 **(90)**
Total	4697 **(480)**	42 **(42)**	1252 **(261)**	13 **(13)**	4014 **(1007)**
SNP density	12.5	11.1	13.3	4.8	15.6
MAF <5%	3615	32	992	13	3198

Table 3: N_{ST} genetic distance measure between study populations

A: NST Genetic Distance between study populations among ORs

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	1.51E-03	3.35E-02	6.82E-02	4.32E-02	7.83E-02	4.82E-02	3.18E-02	2.90E-02	3.19E-02	1.16E-01	1.17E-01	9.03E-02	8.62E-02	1.51E-01	1.79E-01	2.13E-01	2.47E-01
RYK	1.51E-03	0.00E+00	2.67E-02	6.74E-02	4.14E-02	7.24E-02	5.03E-02	3.32E-02	3.41E-02	3.60E-02	9.48E-02	1.03E-01	7.71E-02	7.31E-02	1.24E-01	1.48E-01	1.82E-01	2.17E-01
ANU	3.35E-02	2.67E-02	0.00E+00	9.40E-02	7.63E-02	1.03E-01	8.43E-02	6.60E-02	5.93E-02	7.36E-02	1.19E-01	1.24E-01	1.02E-01	9.25E-02	1.59E-01	1.88E-01	2.18E-01	2.55E-01
AET	6.82E-02	6.74E-02	9.40E-02	0.00E+00	3.54E-02	5.72E-02	5.51E-02	5.45E-02	4.90E-02	4.09E-02	1.29E-01	1.24E-01	1.07E-01	1.03E-01	1.52E-01	1.84E-01	2.04E-01	2.37E-01
AGT	4.32E-02	4.14E-02	7.63E-02	3.54E-02	0.00E+00	5.84E-02	3.81E-02	3.40E-02	3.04E-02	2.30E-02	1.25E-01	1.16E-01	9.72E-02	9.14E-02	1.38E-01	1.61E-01	1.85E-01	2.28E-01
BTK	7.83E-02	7.24E-02	1.03E-01	5.72E-02	5.84E-02	0.00E+00	5.96E-02	6.30E-02	5.95E-02	4.65E-02	1.36E-01	1.39E-01	1.21E-01	1.16E-01	1.53E-01	1.77E-01	1.97E-01	2.43E-01
MNW	4.82E-02	5.03E-02	8.43E-02	5.51E-02	3.81E-02	5.96E-02	0.00E+00	1.98E-02	2.14E-02	2.22E-02	1.32E-01	1.35E-01	1.17E-01	1.05E-01	1.58E-01	1.80E-01	2.04E-01	2.42E-01
MNB	3.18E-02	3.32E-02	6.60E-02	5.45E-02	3.40E-02	6.30E-02	1.98E-02	0.00E+00	2.14E-02	1.51E-02	1.21E-01	1.17E-01	1.04E-01	9.54E-02	1.44E-01	1.68E-01	1.89E-01	2.27E-01
TLG	2.90E-02	3.41E-02	5.93E-02	4.90E-02	3.04E-02	5.95E-02	2.14E-02	2.14E-02	0.00E+00	3.98E-03	1.21E-01	1.23E-01	1.01E-01	9.68E-02	1.57E-01	1.81E-01	2.06E-01	2.47E-01
VSN	3.19E-02	3.60E-02	7.36E-02	4.09E-02	2.30E-02	4.65E-02	2.22E-02	1.51E-02	3.98E-03	0.00E+00	1.27E-01	1.27E-01	1.09E-01	1.07E-01	1.55E-01	1.80E-01	2.01E-01	2.41E-01
DNE	1.16E-01	9.48E-02	1.19E-01	1.29E-01	1.25E-01	1.36E-01	1.32E-01	1.21E-01	1.21E-01	1.27E-01	0.00E+00	1.31E-02	3.30E-04	2.05E-02	8.20E-02	1.16E-01	1.53E-01	2.03E-01
IRS	1.17E-01	1.03E-01	1.24E-01	1.24E-01	1.16E-01	1.39E-01	1.35E-01	1.17E-01	1.23E-01	1.27E-01	1.31E-02	0.00E+00	2.08E-02	1.46E-02	9.05E-02	1.25E-01	1.48E-01	1.98E-01
RUS	9.03E-02	7.71E-02	1.02E-01	1.07E-01	9.72E-02	1.21E-01	1.17E-01	1.04E-01	1.01E-01	1.09E-01	3.30E-04	2.08E-02	0.00E+00	1.76E-02	8.45E-02	1.17E-01	1.52E-01	2.02E-01
ADG	8.62E-02	7.31E-02	9.25E-02	1.03E-01	9.14E-02	1.16E-01	1.05E-01	9.54E-02	9.68E-02	1.07E-01	2.05E-02	1.46E-02	1.76E-02	0.00E+00	8.76E-02	1.19E-01	1.42E-01	1.85E-01
CHG	1.51E-01	1.24E-01	1.59E-01	1.52E-01	1.38E-01	1.53E-01	1.58E-01	1.44E-01	1.57E-01	1.55E-01	8.20E-02	9.05E-02	8.45E-02	8.76E-02	0.00E+00	7.99E-03	4.19E-02	7.76E-02
HSA	1.79E-01	1.48E-01	1.88E-01	1.84E-01	1.61E-01	1.77E-01	1.80E-01	1.68E-01	1.81E-01	1.80E-01	1.16E-01	1.25E-01	1.17E-01	1.19E-01	7.99E-03	0.00E+00	3.30E-02	6.97E-02
BPG	2.13E-01	1.82E-01	2.18E-01	2.04E-01	1.85E-01	1.97E-01	2.04E-01	1.89E-01	2.06E-01	2.01E-01	1.53E-01	1.48E-01	1.52E-01	1.42E-01	4.19E-02	3.30E-02	0.00E+00	6.01E-02
MPG	2.47E-01	2.17E-01	2.55E-01	2.37E-01	2.28E-01	2.43E-01	2.42E-01	2.27E-01	2.47E-01	2.41E-01	2.03E-01	1.98E-01	2.02E-01	1.85E-01	7.76E-02	6.97E-02	6.01E-02	0.00E+00

B: NST Genetic Distance between study populations among neutral references

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	1.55E-03	4.26E-03	5.24E-03	4.28E-03	4.13E-03	4.73E-03	2.61E-03	1.81E-03	3.04E-03	7.90E-03	8.15E-03	7.51E-03	6.89E-03	1.18E-02	1.40E-02	1.67E-02	2.04E-02
RYK	1.55E-03	0.00E+00	4.21E-03	5.62E-03	4.35E-03	4.59E-03	5.36E-03	3.55E-03	2.89E-03	3.83E-03	7.26E-03	7.76E-03	7.30E-03	6.70E-03	1.09E-02	1.31E-02	1.55E-02	1.95E-02
ANU	4.26E-03	4.21E-03	0.00E+00	6.91E-03	6.80E-03	8.00E-03	7.75E-03	5.81E-03	5.18E-03	6.18E-03	1.01E-02	9.42E-03	9.45E-03	9.78E-03	1.28E-02	1.55E-02	1.77E-02	2.13E-02
AET	5.24E-03	5.62E-03	6.91E-03	0.00E+00	4.11E-03	6.12E-03	6.26E-03	4.77E-03	4.29E-03	5.26E-03	9.08E-03	8.22E-03	7.68E-03	9.01E-03	1.16E-02	1.32E-02	1.53E-02	2.00E-02
AGT	4.28E-03	4.35E-03	6.80E-03	4.11E-03	0.00E+00	4.77E-03	5.73E-03	3.86E-03	3.92E-03	4.54E-03	9.05E-03	9.15E-03	7.96E-03	9.09E-03	1.14E-02	1.30E-02	1.47E-02	2.02E-02
BTK	4.13E-03	4.59E-03	8.00E-03	6.12E-03	4.77E-03	0.00E+00	5.32E-03	4.03E-03	3.55E-03	4.00E-03	9.00E-03	9.15E-03	8.67E-03	7.96E-03	1.33E-02	1.52E-02	1.82E-02	2.25E-02
MNW	4.73E-03	5.36E-03	7.75E-03	6.26E-03	5.73E-03	5.32E-03	0.00E+00	3.66E-03	3.97E-03	4.86E-03	1.10E-02	1.02E-02	1.02E-02	1.03E-02	1.38E-02	1.67E-02	1.92E-02	2.36E-02
MNB	2.61E-03	3.55E-03	5.81E-03	4.77E-03	3.86E-03	4.03E-03	3.66E-03	0.00E+00	1.94E-03	3.49E-03	8.19E-03	7.80E-03	6.57E-03	7.46E-03	1.28E-02	1.55E-02	1.79E-02	2.19E-02
TLG	1.81E-03	2.89E-03	5.18E-03	4.29E-03	3.92E-03	3.55E-03	3.97E-03	1.94E-03	0.00E+00	2.40E-03	8.52E-03	8.24E-03	7.72E-03	7.32E-03	1.20E-02	1.41E-02	1.66E-02	2.05E-02
VSN	3.04E-03	3.83E-03	6.18E-03	5.26E-03	4.54E-03	4.00E-03	4.86E-03	3.49E-03	2.40E-03	0.00E+00	9.46E-03	9.40E-03	9.20E-03	8.97E-03	1.21E-02	1.35E-02	1.61E-02	2.02E-02
DNE	7.90E-03	7.26E-03	1.01E-02	9.08E-03	9.05E-03	9.00E-03	1.10E-02	8.19E-03	8.52E-03	9.46E-03	0.00E+00	2.79E-03	2.61E-03	3.06E-03	9.03E-03	1.22E-02	1.50E-02	1.81E-02
IRS	8.15E-03	7.76E-03	9.42E-03	8.22E-03	9.15E-03	9.15E-03	1.02E-02	7.80E-03	8.24E-03	9.40E-03	2.79E-03	0.00E+00	3.07E-03	3.46E-03	9.16E-03	1.27E-02	1.54E-02	1.84E-02
RUS	7.51E-03	7.30E-03	9.45E-03	7.68E-03	7.96E-03	8.67E-03	1.02E-02	6.57E-03	7.72E-03	9.20E-03	2.61E-03	3.07E-03	0.00E+00	3.52E-03	9.96E-03	1.28E-02	1.52E-02	1.89E-02
ADG	6.89E-03	6.70E-03	9.78E-03	9.01E-03	9.09E-03	7.96E-03	1.03E-02	7.46E-03	7.32E-03	8.97E-03	3.06E-03	3.46E-03	3.52E-03	0.00E+00	9.63E-03	1.26E-02	1.56E-02	1.91E-02
CHG	1.18E-02	1.09E-02	1.28E-02	1.16E-02	1.14E-02	1.33E-02	1.38E-02	1.28E-02	1.20E-02	1.21E-02	9.03E-03	9.16E-03	9.96E-03	9.63E-03	0.00E+00	3.73E-03	6.22E-03	8.71E-03
HSA	1.40E-02	1.31E-02	1.55E-02	1.32E-02	1.30E-02	1.52E-02	1.67E-02	1.55E-02	1.41E-02	1.35E-02	1.22E-02	1.27E-02	1.28E-02	1.26E-02	3.73E-03	0.00E+00	5.04E-03	7.79E-03
BPG	1.67E-02	1.55E-02	1.77E-02	1.53E-02	1.47E-02	1.82E-02	1.92E-02	1.79E-02	1.66E-02	1.61E-02	1.50E-02	1.54E-02	1.52E-02	1.56E-02	6.22E-03	5.04E-03	0.00E+00	7.38E-03
MPG	2.04E-02	1.95E-02	2.13E-02	2.00E-02	2.02E-02	2.25E-02	2.36E-02	2.19E-02	2.05E-02	2.02E-02	1.81E-02	1.84E-02	1.89E-02	1.91E-02	8.71E-03	7.79E-03	7.38E-03	0.00E+00

Table 4: d_{xy} genetic distance measure between study populations

A: dxy genetic Distance between study populations among ORs

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	6.39E-02	6.34E-02	6.91E-02	6.65E-02	6.71E-02	6.52E-02	6.51E-02	6.33E-02	6.40E-02	7.42E-02	7.65E-02	7.24E-02	7.43E-02	8.83E-02	9.52E-02	9.95E-02	9.73E-02
RYK	6.39E-02	0.00E+00	6.40E-02	6.99E-02	6.74E-02	6.76E-02	6.65E-02	6.64E-02	6.47E-02	6.53E-02	7.40E-02	7.68E-02	7.28E-02	7.47E-02	8.85E-02	9.53E-02	1.00E-01	9.73E-02
ANU	6.34E-02	6.40E-02	0.00E+00	6.94E-02	6.71E-02	6.71E-02	6.62E-02	6.59E-02	6.37E-02	6.52E-02	7.28E-02	7.54E-02	7.19E-02	7.31E-02	8.78E-02	9.53E-02	9.96E-02	9.74E-02
AET	6.91E-02	6.99E-02	6.94E-02	0.00E+00	6.71E-02	6.73E-02	6.74E-02	6.81E-02	6.61E-02	6.62E-02	7.69E-02	7.82E-02	7.50E-02	7.69E-02	9.01E-02	9.73E-02	1.00E-01	9.76E-02
AGT	6.65E-02	6.74E-02	6.71E-02	6.71E-02	0.00E+00	6.62E-02	6.54E-02	6.59E-02	6.41E-02	6.42E-02	7.61E-02	7.74E-02	7.38E-02	7.58E-02	8.96E-02	9.64E-02	1.00E-01	9.82E-02
BTK	6.71E-02	6.76E-02	6.71E-02	6.73E-02	6.62E-02	0.00E+00	6.53E-02	6.61E-02	6.44E-02	6.41E-02	7.53E-02	7.76E-02	7.38E-02	7.61E-02	8.90E-02	9.58E-02	9.96E-02	9.78E-02
MNW	6.52E-02	6.65E-02	6.62E-02	6.74E-02	6.54E-02	6.53E-02	0.00E+00	6.37E-02	6.23E-02	6.31E-02	7.51E-02	7.72E-02	7.37E-02	7.49E-02	8.93E-02	9.61E-02	9.99E-02	9.73E-02
MNB	6.51E-02	6.64E-02	6.59E-02	6.81E-02	6.59E-02	6.61E-02	6.37E-02	0.00E+00	6.31E-02	6.33E-02	7.51E-02	7.69E-02	7.38E-02	7.54E-02	8.87E-02	9.54E-02	9.88E-02	9.64E-02
TLG	6.33E-02	6.47E-02	6.37E-02	6.61E-02	6.41E-02	6.44E-02	6.23E-02	6.31E-02	0.00E+00	6.10E-02	7.34E-02	7.57E-02	7.19E-02	7.38E-02	8.86E-02	9.55E-02	9.94E-02	9.73E-02
VSN	6.40E-02	6.53E-02	6.52E-02	6.62E-02	6.42E-02	6.41E-02	6.31E-02	6.33E-02	6.10E-02	0.00E+00	7.48E-02	7.70E-02	7.32E-02	7.55E-02	8.94E-02	9.64E-02	1.00E-01	9.78E-02
DNE	7.42E-02	7.40E-02	7.28E-02	7.69E-02	7.61E-02	7.53E-02	7.51E-02	7.51E-02	7.34E-02	7.48E-02	0.00E+00	7.27E-02	7.04E-02	7.29E-02	8.69E-02	9.47E-02	9.95E-02	9.85E-02
IRS	7.65E-02	7.68E-02	7.54E-02	7.82E-02	7.74E-02	7.76E-02	7.72E-02	7.69E-02	7.57E-02	7.70E-02	7.27E-02	0.00E+00	7.33E-02	7.43E-02	8.98E-02	9.78E-02	1.01E-01	1.00E-01
RUS	7.24E-02	7.28E-02	7.19E-02	7.50E-02	7.38E-02	7.38E-02	7.37E-02	7.38E-02	7.19E-02	7.32E-02	7.04E-02	7.33E-02	0.00E+00	7.30E-02	8.75E-02	9.50E-02	9.95E-02	9.82E-02
ADG	7.43E-02	7.47E-02	7.31E-02	7.69E-02	7.58E-02	7.61E-02	7.49E-02	7.54E-02	7.38E-02	7.55E-02	7.29E-02	7.43E-02	7.30E-02	0.00E+00	8.94E-02	9.70E-02	1.00E-01	9.84E-02
CHG	8.83E-02	8.85E-02	8.78E-02	9.01E-02	8.96E-02	8.90E-02	8.93E-02	8.87E-02	8.86E-02	8.94E-02	8.69E-02	8.98E-02	8.75E-02	8.94E-02	0.00E+00	9.46E-02	9.88E-02	9.65E-02
HSA	9.52E-02	9.53E-02	9.53E-02	9.73E-02	9.64E-02	9.58E-02	9.61E-02	9.54E-02	9.55E-02	9.64E-02	9.47E-02	9.78E-02	9.50E-02	9.70E-02	9.46E-02	0.00E+00	1.01E-01	9.96E-02
BPG	9.95E-02	1.00E-01	9.96E-02	1.00E-01	1.00E-01	9.96E-02	9.99E-02	9.88E-02	9.94E-02	1.00E-01	9.95E-02	1.01E-01	9.95E-02	1.00E-01	9.88E-02	1.01E-01	0.00E+00	9.96E-02
MPG	9.73E-02	9.73E-02	9.74E-02	9.76E-02	9.82E-02	9.78E-02	9.73E-02	9.64E-02	9.73E-02	9.78E-02	9.85E-02	1.00E-01	9.82E-02	9.84E-02	9.65E-02	9.96E-02	9.96E-02	0.00E+00

B: dxy genetic Distance between study populations among neutral references

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	5.96E-02	6.23E-02	6.66E-02	6.50E-02	6.41E-02	6.34E-02	6.24E-02	6.16E-02	6.36E-02	6.80E-02	6.88E-02	6.80E-02	6.81E-02	7.79E-02	8.17E-02	8.36E-02	8.61E-02
RYK	5.96E-02	0.00E+00	6.04E-02	6.51E-02	6.32E-02	6.26E-02	6.21E-02	6.10E-02	6.06E-02	6.19E-02	6.67E-02	6.75E-02	6.69E-02	6.65E-02	7.73E-02	8.09E-02	8.29E-02	8.53E-02
ANU	6.23E-02	6.04E-02	0.00E+00	6.58E-02	6.53E-02	6.55E-02	6.46E-02	6.31E-02	6.24E-02	6.43E-02	6.79E-02	6.76E-02	6.73E-02	6.85E-02	7.69E-02	8.16E-02	8.29E-02	8.48E-02
AET	6.66E-02	6.51E-02	6.58E-02	0.00E+00	6.55E-02	6.69E-02	6.57E-02	6.56E-02	6.49E-02	6.67E-02	6.90E-02	6.88E-02	6.83E-02	6.99E-02	7.87E-02	8.28E-02	8.42E-02	8.71E-02
AGT	6.50E-02	6.32E-02	6.53E-02	6.55E-02	0.00E+00	6.47E-02	6.46E-02	6.42E-02	6.34E-02	6.50E-02	6.87E-02	6.93E-02	6.81E-02	6.92E-02	7.80E-02	8.12E-02	8.25E-02	8.64E-02
BTK	6.41E-02	6.26E-02	6.55E-02	6.69E-02	6.47E-02	0.00E+00	6.42E-02	6.33E-02	6.25E-02	6.37E-02	6.89E-02	6.91E-02	6.89E-02	6.84E-02	7.95E-02	8.30E-02	8.52E-02	8.85E-02
MNW	6.34E-02	6.21E-02	6.46E-02	6.57E-02	6.46E-02	6.42E-02	0.00E+00	6.11E-02	6.17E-02	6.37E-02	6.94E-02	6.98E-02	6.92E-02	6.97E-02	7.96E-02	8.40E-02	8.56E-02	8.89E-02
MNB	6.24E-02	6.10E-02	6.31E-02	6.56E-02	6.42E-02	6.33E-02	6.11E-02	0.00E+00	6.07E-02	6.29E-02	6.79E-02	6.83E-02	6.73E-02	6.84E-02	7.92E-02	8.36E-02	8.52E-02	8.77E-02
TLG	6.16E-02	6.06E-02	6.24E-02	6.49E-02	6.34E-02	6.25E-02	6.17E-02	6.07E-02	0.00E+00	6.20E-02	6.71E-02	6.76E-02	6.69E-02	6.74E-02	7.72E-02	8.14E-02	8.32E-02	8.59E-02
VSN	6.36E-02	6.19E-02	6.43E-02	6.67E-02	6.50E-02	6.37E-02	6.37E-02	6.29E-02	6.20E-02	0.00E+00	6.89E-02	6.95E-02	6.90E-02	6.91E-02	7.80E-02	8.16E-02	8.37E-02	8.63E-02
DNE	6.80E-02	6.67E-02	6.79E-02	6.90E-02	6.87E-02	6.89E-02	6.94E-02	6.79E-02	6.71E-02	6.89E-02	0.00E+00	6.26E-02	6.26E-02	6.34E-02	7.56E-02	8.04E-02	8.28E-02	8.45E-02
IRS	6.88E-02	6.75E-02	6.76E-02	6.88E-02	6.93E-02	6.91E-02	6.98E-02	6.83E-02	6.76E-02	6.95E-02	6.26E-02	0.00E+00	6.29E-02	6.39E-02	7.56E-02	8.10E-02	8.33E-02	8.50E-02
RUS	6.80E-02	6.69E-02	6.73E-02	6.83E-02	6.81E-02	6.89E-02	6.92E-02	6.73E-02	6.69E-02	6.90E-02	6.26E-02	6.29E-02	0.00E+00	6.42E-02	7.60E-02	8.06E-02	8.28E-02	8.50E-02
ADG	6.81E-02	6.65E-02	6.85E-02	6.99E-02	6.92E-02	6.84E-02	6.97E-02	6.84E-02	6.74E-02	6.91E-02	6.34E-02	6.39E-02	6.42E-02	0.00E+00	7.63E-02	8.05E-02	8.33E-02	8.55E-02
CHG	7.79E-02	7.73E-02	7.69E-02	7.87E-02	7.80E-02	7.95E-02	7.96E-02	7.92E-02	7.72E-02	7.80E-02	7.56E-02	7.56E-02	7.60E-02	7.63E-02	0.00E+00	7.86E-02	8.14E-02	8.25E-02
HSA	8.17E-02	8.09E-02	8.16E-02	8.28E-02	8.12E-02	8.30E-02	8.40E-02	8.36E-02	8.14E-02	8.16E-02	8.04E-02	8.10E-02	8.06E-02	8.05E-02	7.86E-02	0.00E+00	8.20E-02	8.34E-02
BPG	8.36E-02	8.29E-02	8.29E-02	8.42E-02	8.25E-02	8.52E-02	8.56E-02	8.52E-02	8.32E-02	8.37E-02	8.28E-02	8.33E-02	8.28E-02	8.33E-02	8.14E-02	8.20E-02	0.00E+00	8.30E-02
MPG	8.61E-02	8.53E-02	8.48E-02	8.71E-02	8.64E-02	8.85E-02	8.89E-02	8.77E-02	8.59E-02	8.63E-02	8.45E-02	8.50E-02	8.50E-02	8.55E-02	8.25E-02	8.34E-02	8.30E-02	0.00E+00

Table 5: d_A genetic distance measure between study populations

A: dA genetic distance between study populations among ORs

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	1.49E-04	2.17E-03	4.99E-03	2.96E-03	5.10E-03	3.05E-03	2.03E-03	1.93E-03	2.03E-03	7.87E-03	8.25E-03	5.76E-03	6.54E-03	1.18E-02	1.49E-02	1.83E-02	2.23E-02
RYK	1.49E-04	0.00E+00	1.74E-03	4.84E-03	2.88E-03	4.67E-03	3.29E-03	2.29E-03	2.34E-03	2.33E-03	6.63E-03	7.59E-03	5.08E-03	5.87E-03	1.11E-02	1.41E-02	1.78E-02	2.12E-02
ANU	2.17E-03	1.74E-03	0.00E+00	6.88E-03	5.11E-03	6.66E-03	5.59E-03	4.38E-03	3.85E-03	4.73E-03	8.04E-03	8.69E-03	6.73E-03	6.81E-03	1.29E-02	1.66E-02	1.99E-02	2.39E-02
AET	4.99E-03	4.84E-03	6.88E-03	0.00E+00	2.28E-03	4.04E-03	3.89E-03	3.72E-03	3.35E-03	2.88E-03	9.24E-03	8.64E-03	6.95E-03	7.74E-03	1.24E-02	1.57E-02	1.79E-02	2.12E-02
AGT	2.96E-03	2.88E-03	5.11E-03	2.28E-03	0.00E+00	3.54E-03	2.47E-03	2.12E-03	1.93E-03	1.45E-03	9.00E-03	8.48E-03	6.36E-03	7.26E-03	1.25E-02	1.54E-02	1.81E-02	2.24E-02
BTK	5.10E-03	4.67E-03	6.66E-03	4.04E-03	3.54E-03	0.00E+00	3.91E-03	3.86E-03	3.83E-03	2.84E-03	9.71E-03	1.02E-02	7.96E-03	9.08E-03	1.34E-02	1.64E-02	1.91E-02	2.35E-02
MNW	3.05E-03	3.29E-03	5.59E-03	3.89E-03	2.47E-03	3.91E-03	0.00E+00	1.21E-03	1.51E-03	1.67E-03	9.30E-03	9.57E-03	7.55E-03	7.70E-03	1.34E-02	1.64E-02	1.92E-02	2.28E-02
MNB	2.03E-03	2.29E-03	4.38E-03	3.72E-03	2.12E-03	3.86E-03	1.21E-03	0.00E+00	1.42E-03	9.51E-04	8.47E-03	8.45E-03	6.83E-03	7.34E-03	1.19E-02	1.49E-02	1.72E-02	2.11E-02
TLG	1.93E-03	2.34E-03	3.85E-03	3.35E-03	1.93E-03	3.83E-03	1.51E-03	1.42E-03	0.00E+00	3.61E-04	8.43E-03	8.85E-03	6.56E-03	7.35E-03	1.35E-02	1.66E-02	1.95E-02	2.36E-02
VSN	2.03E-03	2.33E-03	4.73E-03	2.88E-03	1.45E-03	2.84E-03	1.67E-03	9.51E-04	3.61E-04	0.00E+00	9.17E-03	9.50E-03	7.25E-03	8.48E-03	1.37E-02	1.69E-02	1.97E-02	2.35E-02
DNE	7.87E-03	6.63E-03	8.04E-03	9.24E-03	9.00E-03	9.71E-03	9.30E-03	8.47E-03	8.43E-03	9.17E-03	0.00E+00	8.73E-04	9.11E-05	1.54E-03	6.92E-03	1.08E-02	1.46E-02	1.99E-02
IRS	8.25E-03	7.59E-03	8.69E-03	8.64E-03	8.48E-03	1.02E-02	9.57E-03	8.45E-03	8.85E-03	9.50E-03	8.73E-04	0.00E+00	1.13E-03	1.02E-03	7.88E-03	1.20E-02	1.48E-02	1.98E-02
RUS	5.76E-03	5.08E-03	6.73E-03	6.95E-03	6.36E-03	7.96E-03	7.55E-03	6.83E-03	6.56E-03	7.25E-03	9.11E-05	1.13E-03	0.00E+00	1.32E-03	7.14E-03	1.08E-02	1.43E-02	1.93E-02
ADG	6.54E-03	5.87E-03	6.81E-03	7.74E-03	7.26E-03	9.08E-03		7.34E-03	7.35E-03	8.48E-03	1.54E-03	1.02E-03	1.32E-03	0.00E+00	7.88E-03	1.17E-02	1.42E-02	1.83E-02
CHG	1.18E-02	1.11E-02	1.29E-02	1.24E-02	1.25E-02	1.34E-02	1.34E-02	1.19E-02	1.35E-02	1.37E-02	6.92E-03	7.88E-03	7.14E-03	7.88E-03	0.00E+00	6.83E-04	3.86E-03	7.80E-03
HSA	1.49E-02	1.41E-02	1.66E-02	1.57E-02	1.54E-02	1.64E-02	1.64E-02	1.49E-02	1.66E-02	1.69E-02	1.08E-02	1.20E-02	1.08E-02	1.17E-02	6.83E-04	0.00E+00	2.72E-03	7.04E-03
BPG	1.83E-02	1.78E-02	1.99E-02	1.79E-02	1.81E-02	1.91E-02	1.92E-02	1.72E-02	1.95E-02	1.97E-02	1.46E-02	1.48E-02	1.43E-02	1.42E-02	3.86E-03	2.72E-03	0.00E+00	6.06E-03
MPG	2.23E-02	2.12E-02	2.39E-02	2.12E-02	2.24E-02	2.35E-02	2.28E-02	2.11E-02	2.36E-02	2.35E-02	1.99E-02	1.98E-02	1.93E-02	1.83E-02	7.80E-03	7.04E-03	6.06E-03	0.00E+00

B: dA genetic distance between study populations among neutral references

	HNS	RYK	ANU	AET	AGT	BTK	MNW	MNB	TLG	VSN	DNE	IRS	RUS	ADG	CHG	HSA	BPG	MPG
HNS	0.00E+00	-3.81E-04	3.41E-03	4.45E-03	2.91E-03	3.06E-03	3.15E-03	2.08E-03	9.82E-04	1.38E-03	6.47E-03	7.28E-03	6.04E-03	5.76E-03	9.16E-03	1.15E-02	1.30E-02	1.67E-02
RYK	-3.81E-04	0.00E+00	2.54E-03	4.01E-03	2.18E-03	2.55E-03	2.84E-03	1.71E-03	1.02E-03	6.76E-04	6.18E-03	7.01E-03	5.97E-03	5.20E-03	9.52E-03	1.17E-02	1.33E-02	1.70E-02
ANU	3.41E-03	2.54E-03	0.00E+00	5.73E-03	5.37E-03	6.50E-03	6.39E-03	4.86E-03	3.91E-03	4.18E-03	8.45E-03	8.22E-03	7.43E-03	8.23E-03	1.03E-02	1.35E-02	1.44E-02	1.75E-02
AET	4.45E-03	4.01E-03	5.73E-03	0.00E+00	2.29E-03	4.70E-03	4.28E-03	4.15E-03	3.17E-03	3.31E-03	6.32E-03	6.16E-03	5.23E-03	6.41E-03	8.84E-03	1.14E-02	1.24E-02	1.66E-02
AGT	2.91E-03	2.18E-03	5.37E-03	2.29E-03	0.00E+00	2.63E-03	3.26E-03	2.77E-03	1.79E-03	1.69E-03	6.16E-03	6.81E-03	5.15E-03	5.82E-03	8.22E-03	9.91E-03	1.09E-02	1.60E-02
BTK	3.06E-03	2.55E-03	6.50E-03	4.70E-03	2.63E-03	0.00E+00	3.88E-03	2.93E-03	1.89E-03	1.38E-03	7.29E-03	7.62E-03	6.92E-03	6.07E-03	1.07E-02	1.27E-02	1.46E-02	1.91E-02
MNW	3.15E-03	2.84E-03	6.39E-03	4.28E-03	3.26E-03	3.88E-03	0.00E+00	1.53E-03	1.91E-03	2.23E-03	8.66E-03	9.03E-03	8.05E-03	8.09E-03	1.16E-02	1.45E-02	1.57E-02	2.03E-02
MNB	2.08E-03	1.71E-03	4.86E-03	4.15E-03	2.77E-03	2.93E-03	1.53E-03	0.00E+00	7.16E-04	1.34E-03	7.09E-03	7.43E-03	5.99E-03	6.71E-03	1.11E-02	1.40E-02	1.52E-02	1.90E-02
TLG	9.82E-04	1.02E-03	3.91E-03	3.17E-03	1.79E-03	1.89E-03	1.91E-03	7.16E-04	0.00E+00	1.51E-04	6.05E-03	6.55E-03	5.43E-03	5.52E-03	8.89E-03	1.16E-02	1.30E-02	1.69E-02
VSN	1.38E-03	6.76E-04	4.18E-03	3.31E-03	1.69E-03	1.38E-03	2.23E-03	1.34E-03	1.51E-04	0.00E+00	6.15E-03	6.76E-03	5.85E-03	5.56E-03	8.01E-03	1.01E-02	1.18E-02	1.57E-02
DNE	6.47E-03	6.18E-03	8.45E-03	6.32E-03	6.16E-03	7.29E-03	8.66E-03	7.09E-03	6.05E-03	6.15E-03	0.00E+00	6.70E-04	1.79E-04	5.91E-04	6.29E-03	9.66E-03	1.17E-02	1.47E-02
IRS	7.28E-03	7.01E-03	8.22E-03	6.16E-03	6.81E-03	7.62E-03	9.03E-03	7.43E-03	6.55E-03	6.76E-03	6.70E-04	0.00E+00	5.54E-04	1.14E-03	6.39E-03	1.04E-02	1.22E-02	1.51E-02
RUS	6.04E-03	5.97E-03	7.43E-03	5.23E-03	5.15E-03	6.92E-03	8.05E-03	5.99E-03	5.43E-03	5.85E-03	1.79E-04	5.54E-04	0.00E+00	9.46E-04	6.34E-03	9.49E-03	1.12E-02	1.47E-02
ADG	5.76E-03	5.20E-03	8.23E-03	6.41E-03	5.82E-03	6.07E-03	8.09E-03	6.71E-03	5.52E-03	5.56E-03	5.91E-04	1.14E-03	9.46E-04	0.00E+00	6.20E-03	8.92E-03	1.14E-02	1.48E-02
CHG	9.16E-03	9.52E-03	1.03E-02	8.84E-03	8.22E-03	1.07E-02	1.16E-02	1.11E-02	8.89E-03	8.01E-03	6.29E-03	6.39E-03	6.34E-03	6.20E-03	0.00E+00	6.71E-04	3.06E-03	5.35E-03
HSA	1.15E-02	1.17E-02	1.35E-02	1.14E-02	9.91E-03	1.27E-02	1.45E-02	1.40E-02	1.16E-02	1.01E-02	9.66E-03	1.04E-02	9.49E-03	8.92E-03	6.71E-04	0.00E+00	2.22E-03	4.82E-03
BPG	1.30E-02	1.33E-02	1.44E-02	1.24E-02	1.09E-02	1.46E-02	1.57E-02	1.52E-02	1.30E-02	1.18E-02	1.17E-02	1.22E-02	1.12E-02	1.14E-02	3.06E-03	2.22E-03	0.00E+00	4.04E-03
MPG	1.67E-02	1.70E-02	1.75E-02	1.66E-02	1.60E-02	1.91E-02	2.03E-02	1.90E-02	1.69E-02	1.57E-02	1.47E-02	1.51E-02	1.47E-02	1.48E-02	5.35E-03	4.82E-03	4.04E-03	0.00E+00

Table 6: Featured OR genes in this study

OR Gene	Ligand	Nasal Expression	Non-nasal Expression	Feature
HsOR1.4.9	-	+	Cerebral Cortex, Testis	• Disrupted in many individuals
HsOR3.3.6	-	+	Esophagus, Tibial Nerve	• Disrupted in many individuals • High Tajima's D
HsOR11.3.63	-	+	Whole Blood, Spinal Cord	• Disrupted in many individuals • Found disrupted in Denisovans and Neanderthals • Deletion allele • High Tajima's *D*
HsOR11.8.18P0	-	-	Ovary, Skeletal Muscle	• Disrupted in all individuals • High frequency of triplication allele
HsOR11.11.21P0	-	-	Tibial Artery, Cerebral Cortex	• Disrupted in all individuals • High frequency of deletion allele
HsOR11.11.63	-	+	Tibial Nerve, Skin	• Intact in all individuals • High frequency of deletion allele
HsOR11.12.5	-	+	Spinal Cord, Skin	• Intact in all individuals • High N_{ST}
HsOR14.1.20	-	-	Tibial Nerve, Esophagus	• Disrupted in many individuals
HsOR14.1.28P0	Isovaleric Acid	+	-	• Intact in many individuals

FIGURES

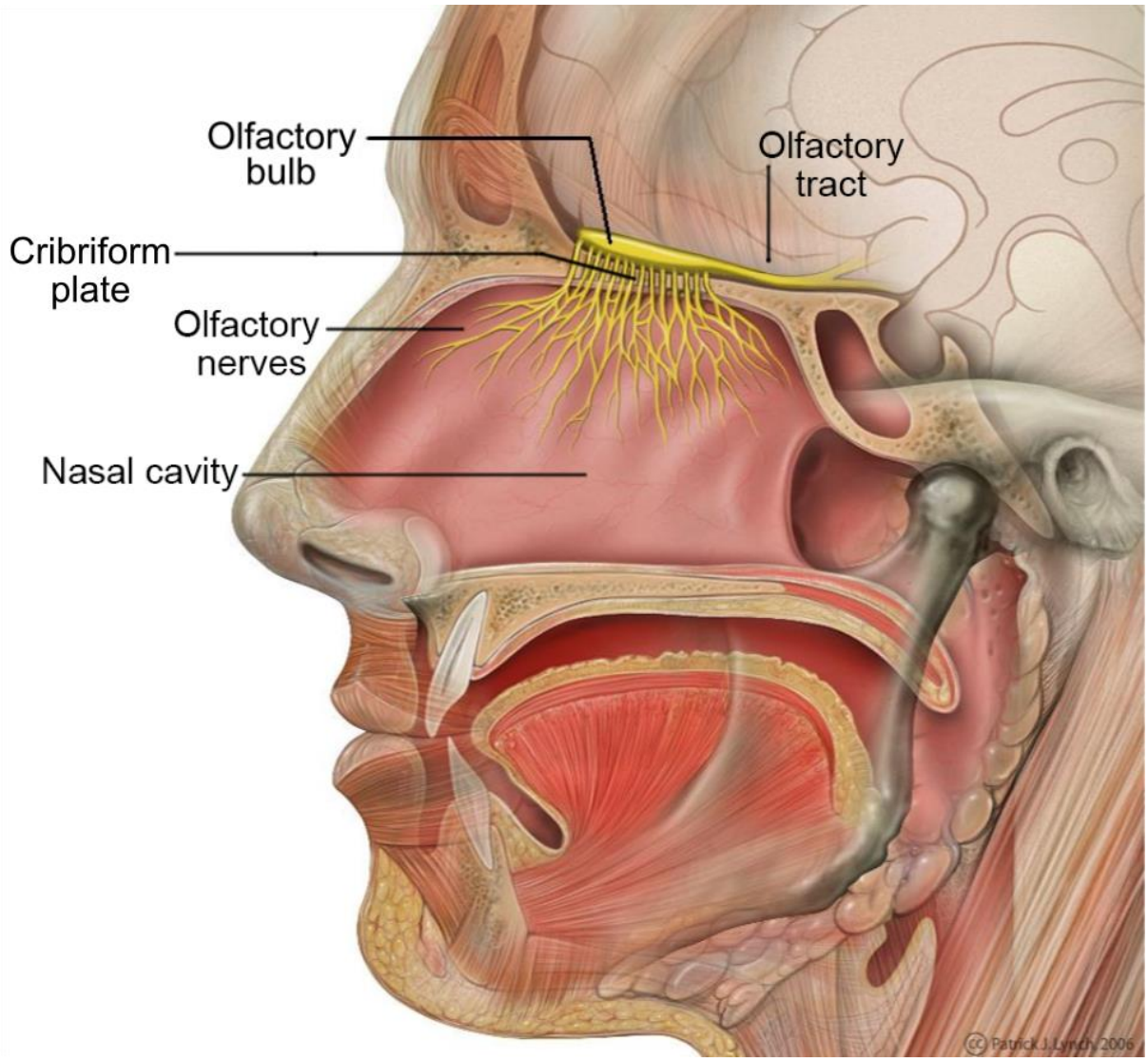

Figure 1: Anatomical representation of human olfactory system. Nasal cavity connecting olfactory bulb in brain via olfactory nerves/olfactory sensory neurons is shown

(Illustration source: Wikimedia commons under CC BY 2.5)

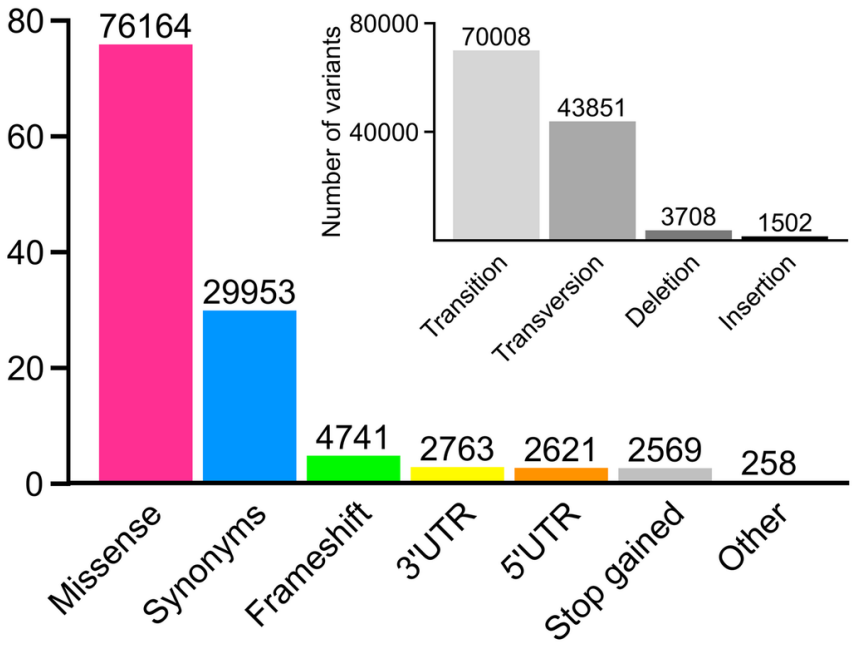

Figure 2: Distribution of variants among structural categories as present in hORMdb. Vertical axis represents frequencies (x10^3) and horizontal axis is representative of variant type. Graph is obtained from Jimenez et al. 2021 under CC BY 4.0 license

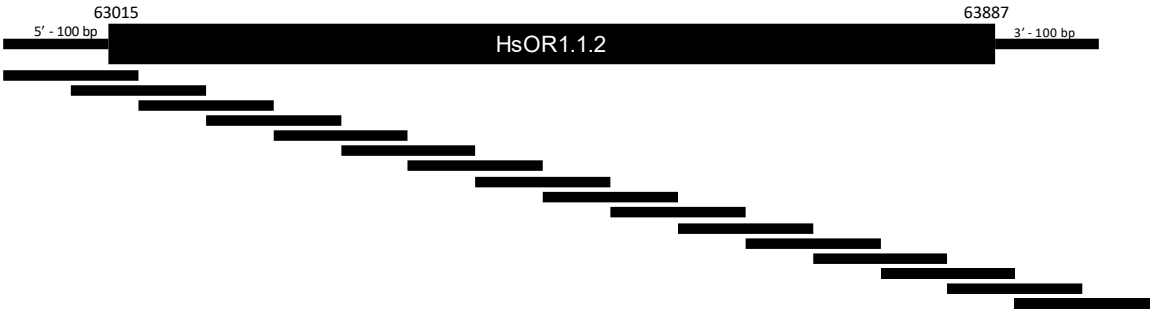

Figure 3: Exemplary schematic illustration of target capture probes of 120 nucleotides each with 60 nucleotides overlapped. HsOR1.1.2 in this illustration is covered by a set of 16 probes covering coding region and 100 – bp flanking sequences

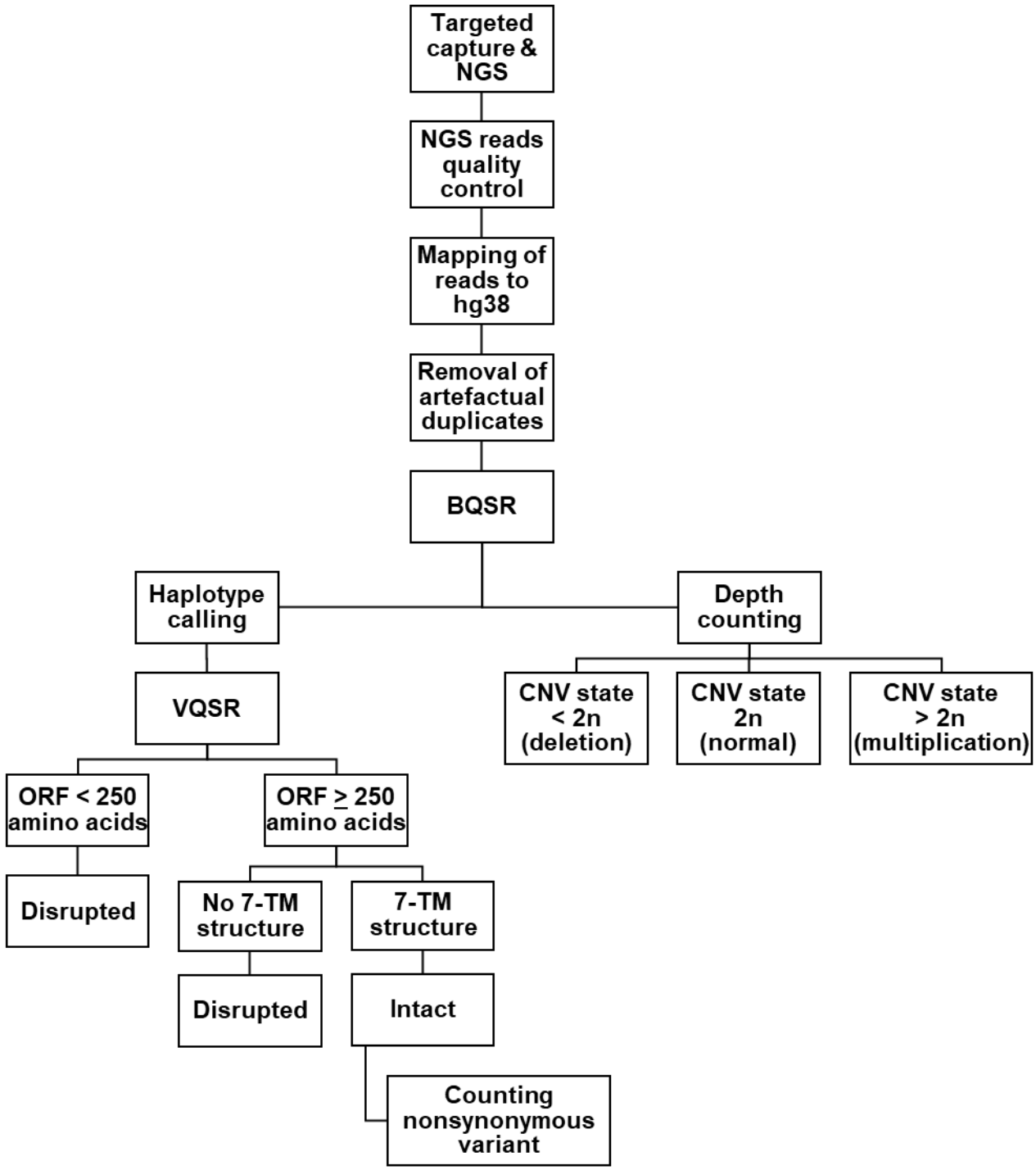

Figure 4: Flowchart of NGS Reads Analysis

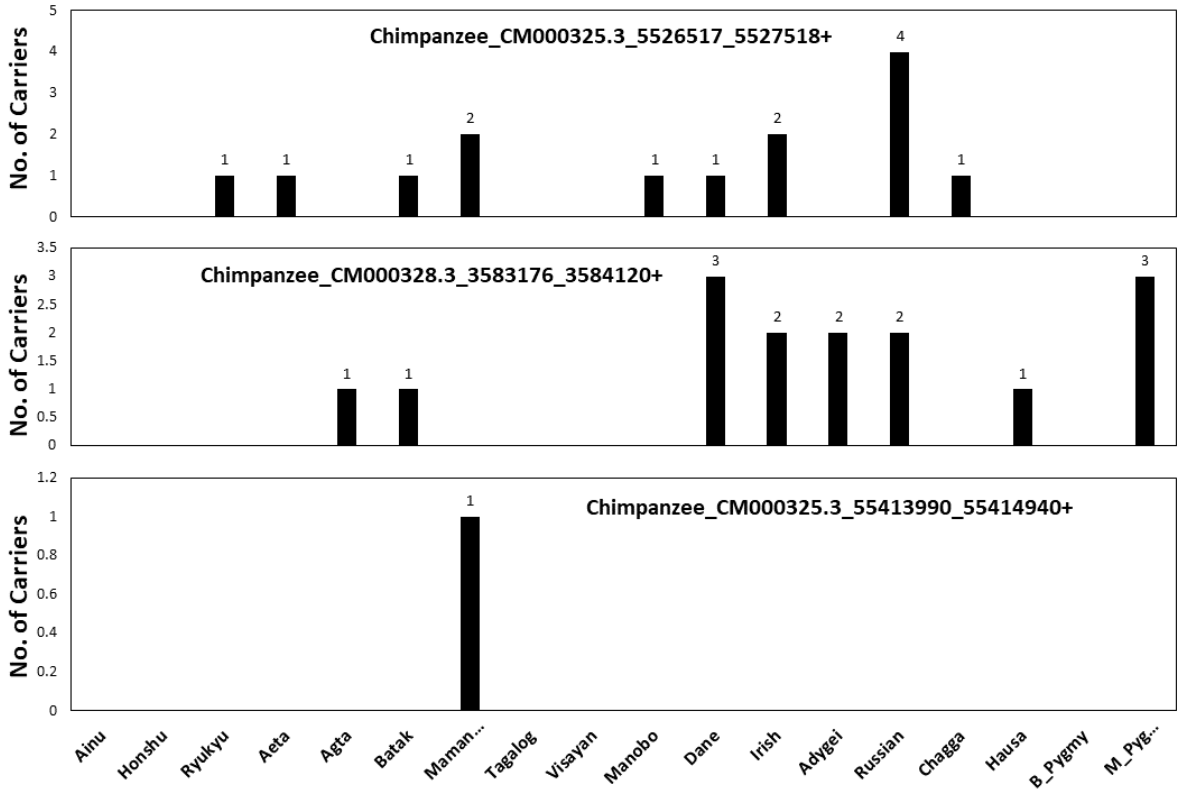

Figure 5: Retrieval frequency of three Pantro OR genes among study populations. Populations are given on horizontal axis while number of carriers in each population are given on vertical axis. No bar is shown if gene was not retrieved in any individual.

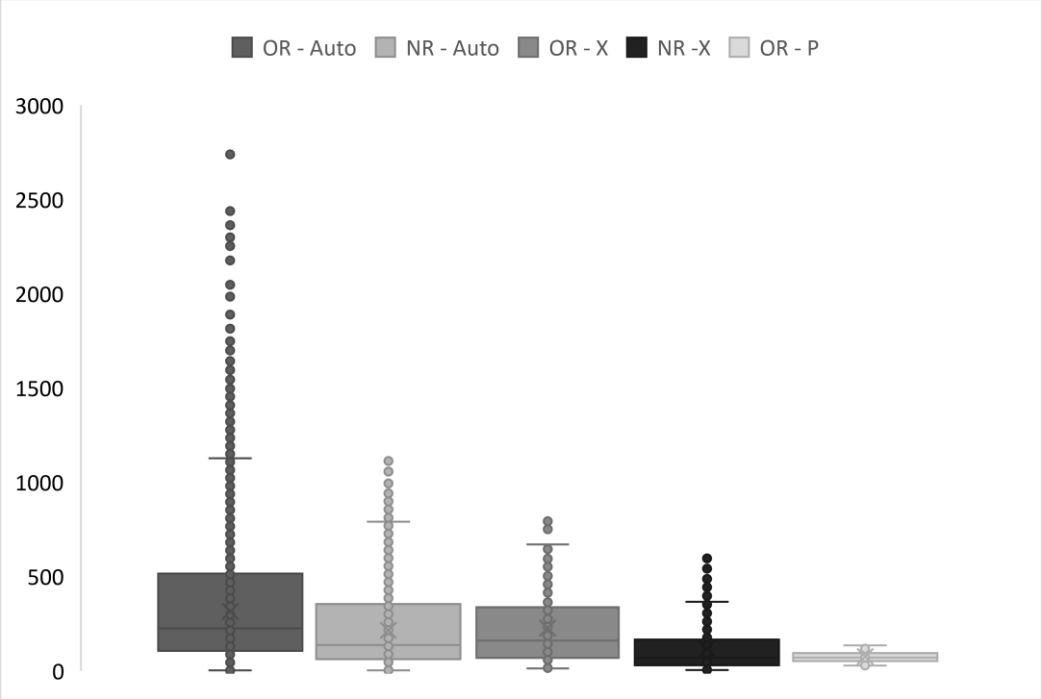

Figure 6: Box plots of sequencing depths for all data throughout genes/regions and throughout carrier individuals in each gene/region category. The data from the targeted capture are categorized into the autosomal OR genes (OR-A), autosomal neutral references (NR-A), an X-chromosomal OR gene (OR-X), X-chromosomal neutral references (NR-X) and three Pantro 3.0 OR gene detected in this study (OR-P).

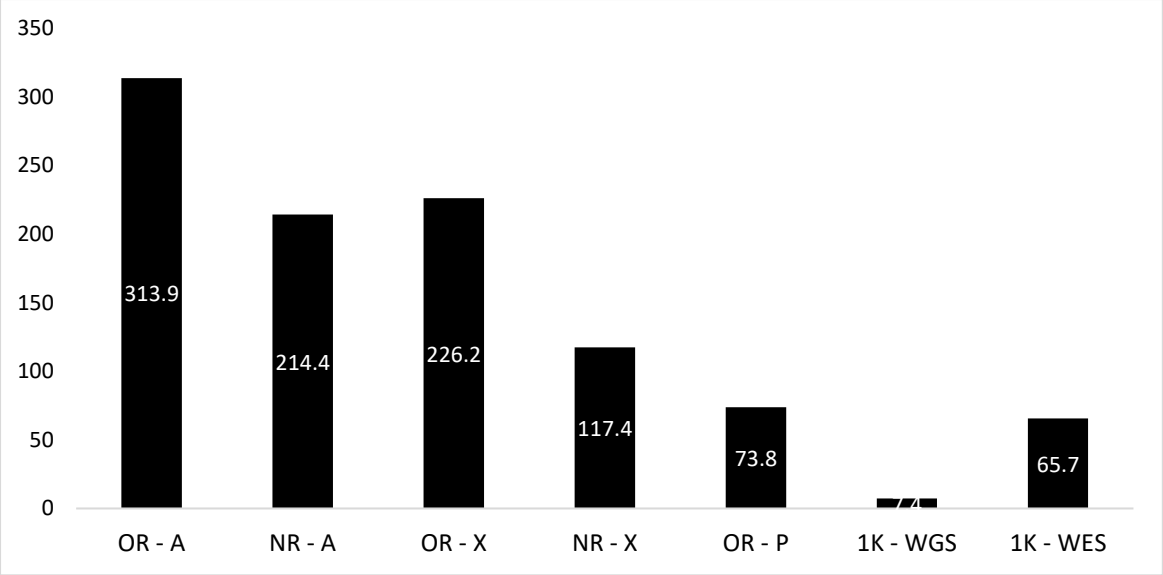

Figure 7: Comparison of mean depths among targeted captures OR-A, NR-A, OR-X, NR-X and OR-P, the whole-genome sequencing in the 1000 Genomes project (1K WGS) and the whole-exome sequencing in the 1000 Genomes project (1K WES).

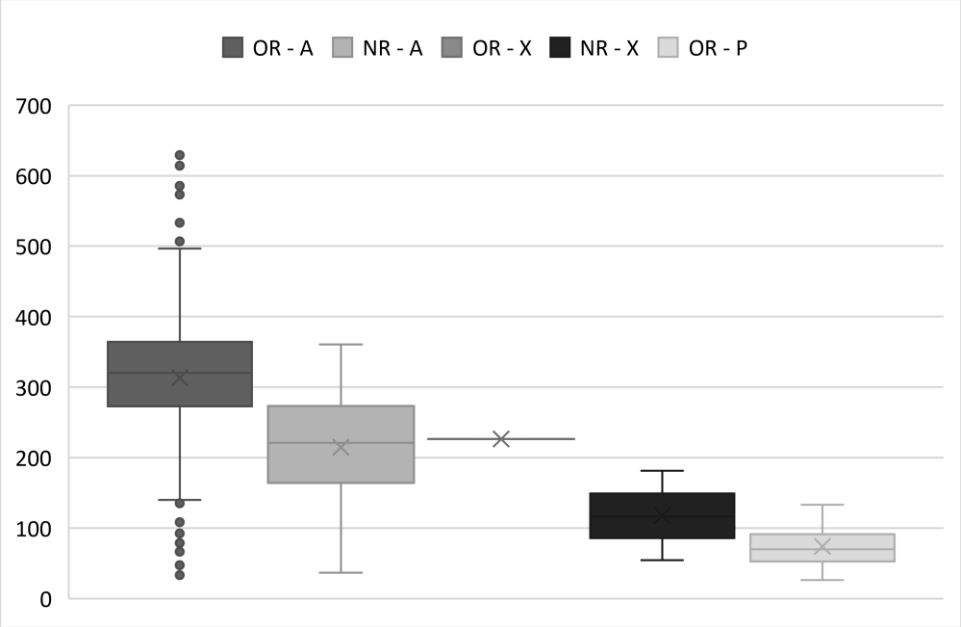

Figure 8: Box plots of sequencing depths averaged throughout carrier individuals for each gene/region in each gene/region category

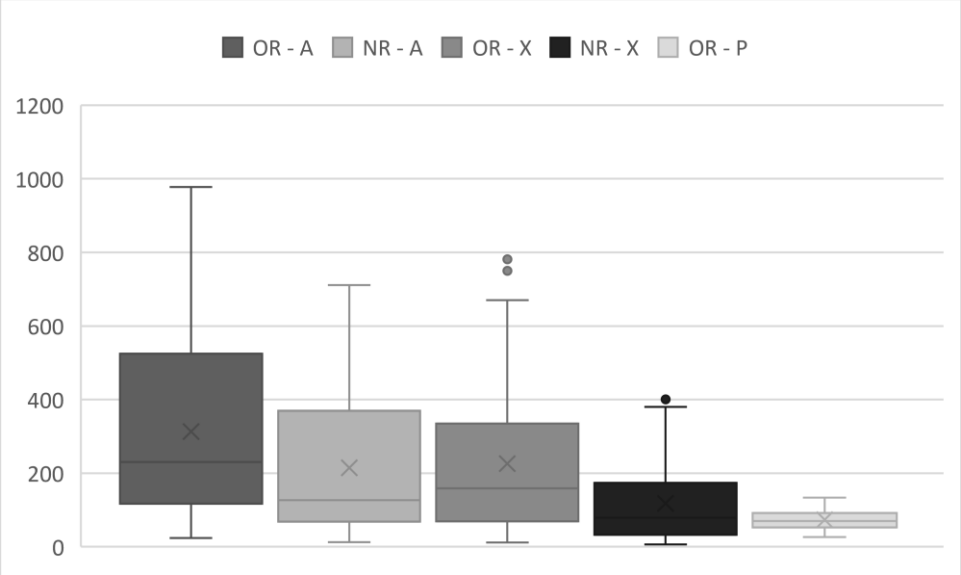

Figure 9: Box plots of sequencing depths averaged throughout genes/regions in each gene/region category for each carrier individual

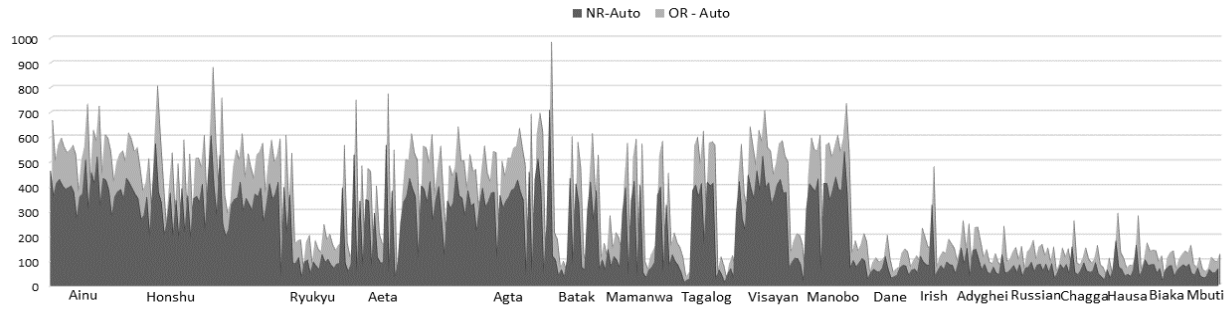

Figure 10: Mean depth of autosomal ORs and neutral references per carrier plotted by population.

Japanese individuals followed by Filipino, European and African individuals are shown.

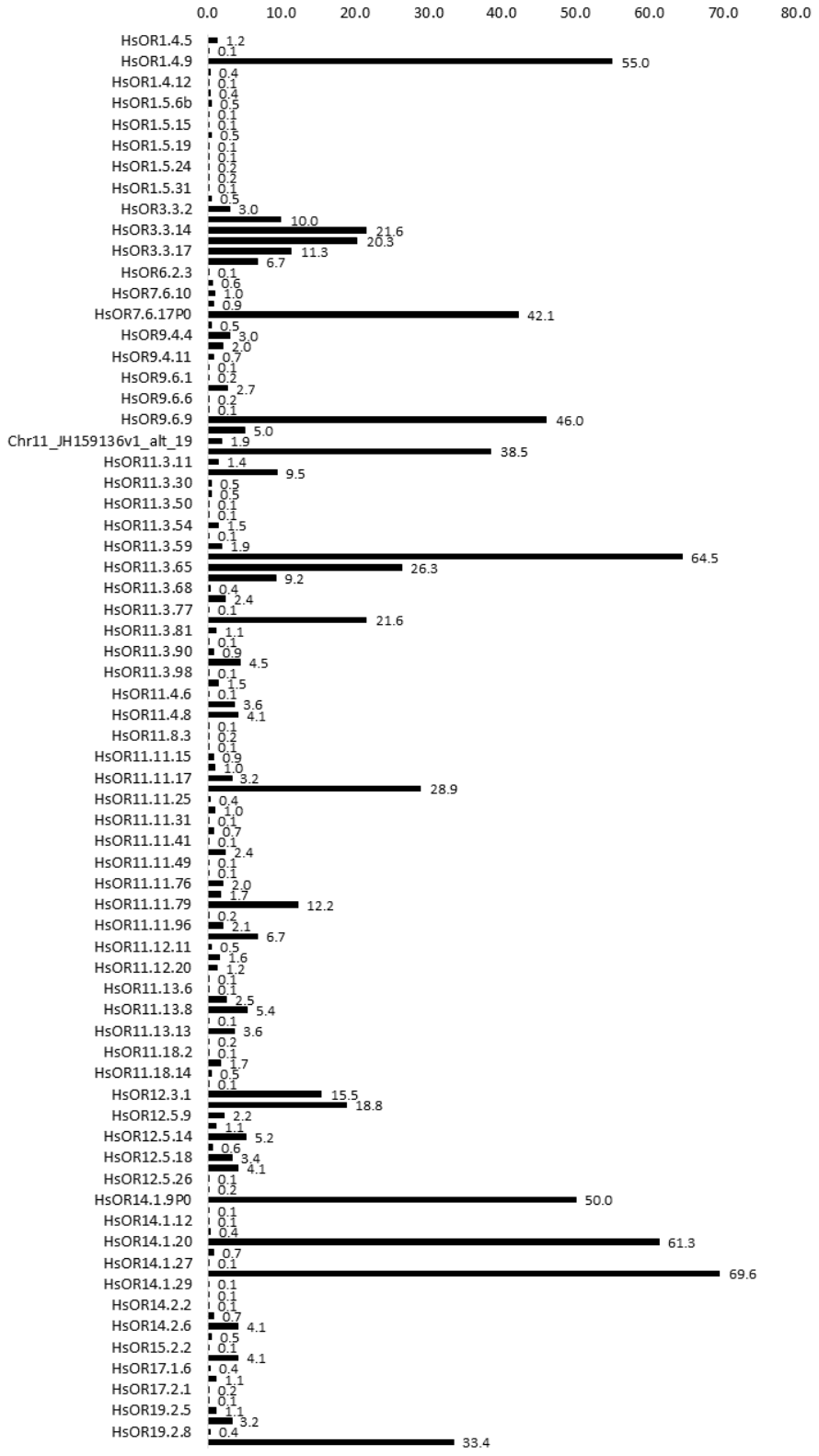

Figure 11: Frequencies of disrupted alleles of the OR segregating pseudogenes in the 401 study individuals. Due to limited space, gene name labels are not complete in figure so disrupted allele frequency of all 134 OR genes are also given in Appendix VI.

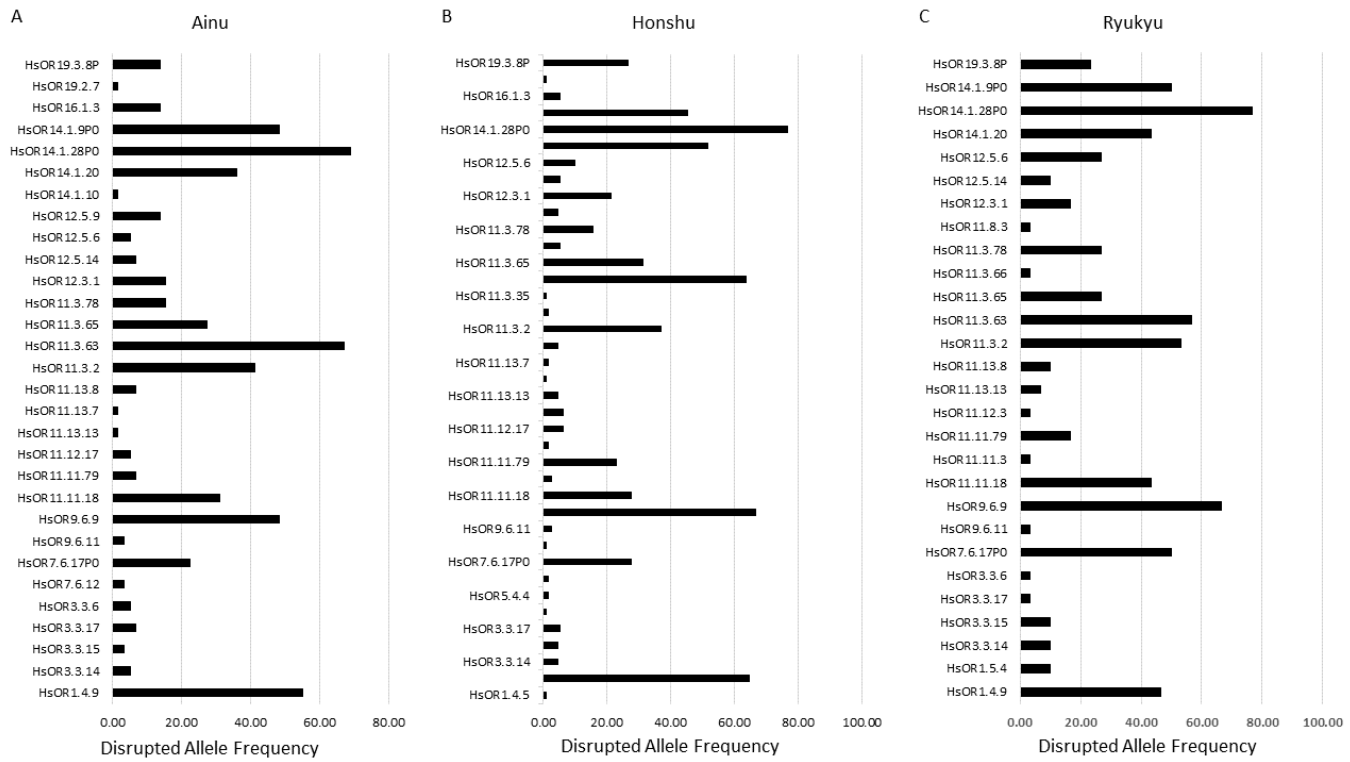

Figure 12: Disrupted Allele frequencies across study individuals of three Japanese populations. Given allele frequencies are in percentage and ordered in chromosomal order from bottom to top. Vertical axis represents gene name label while horizontal axis represents incidence of disrupted allele frequencies.

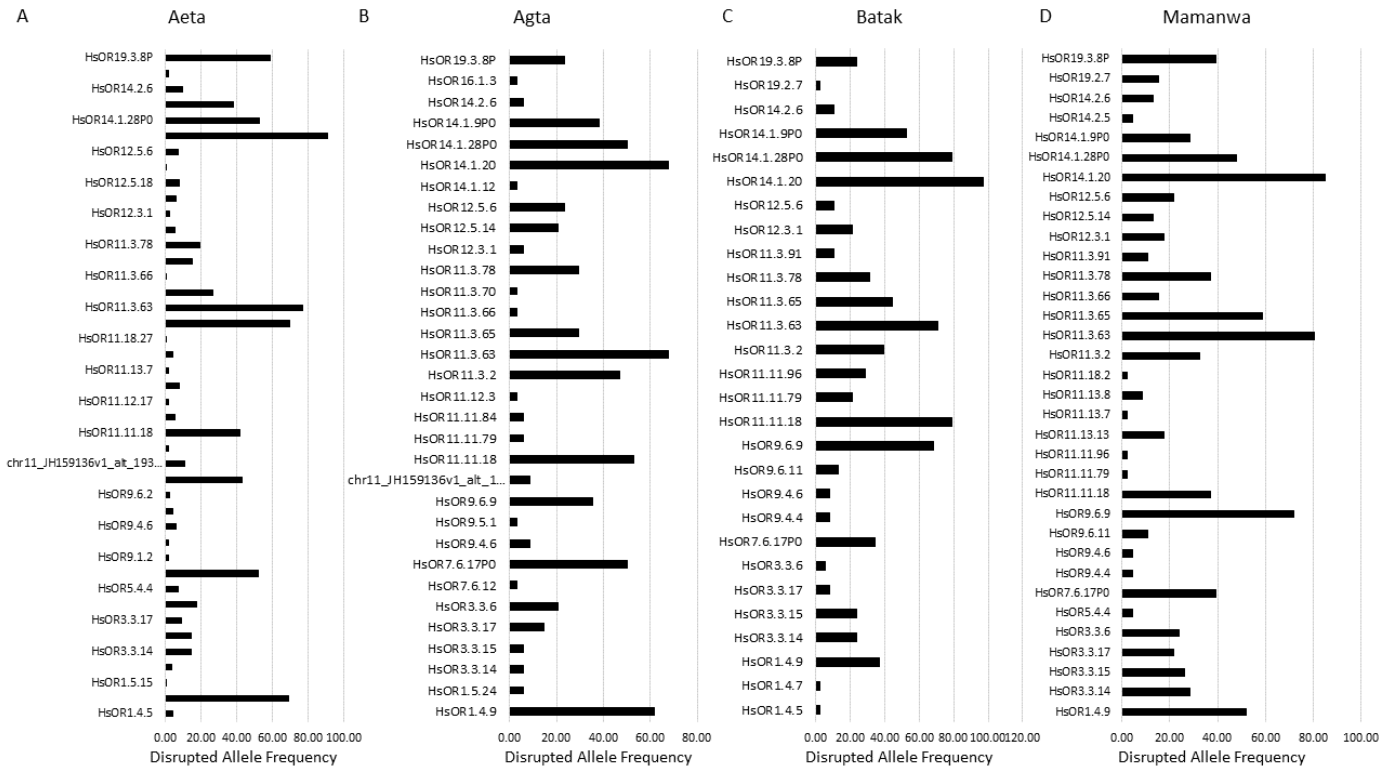

Figure 13: Disrupted Allele frequencies across study individuals of four negrito populations of Philippines. Given allele frequencies are in percentage and ordered in chromosomal order from bottom to top. Vertical axis represents gene name label while horizontal axis represents incidence of disrupted allele frequencies.

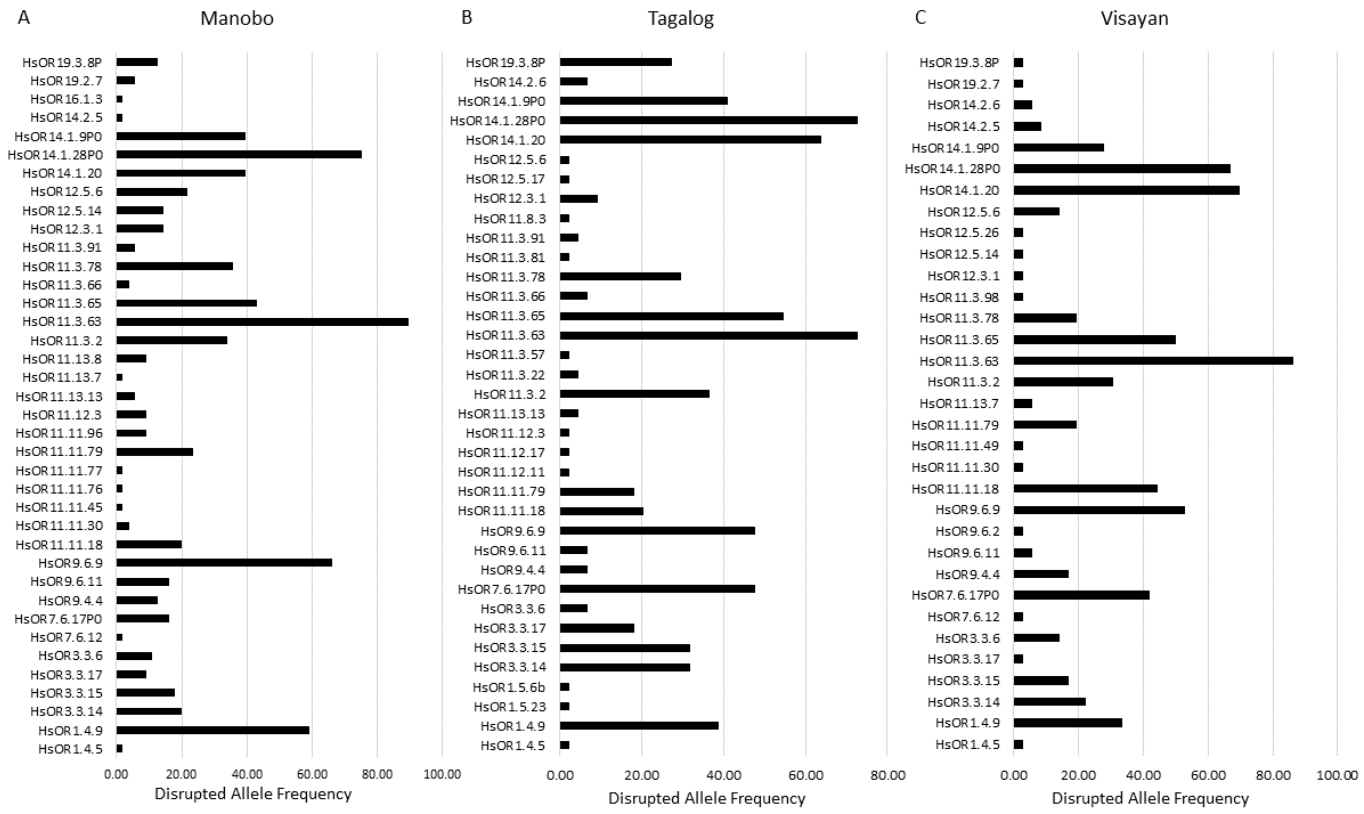

Figure 14: Disrupted Allele frequencies across study individuals of four agricultural populations of Philippines. Given allele frequencies are in percentage and ordered in chromosomal order from bottom to top. Vertical axis represents gene name label while horizontal axis represents incidence of disrupted allele frequencies.

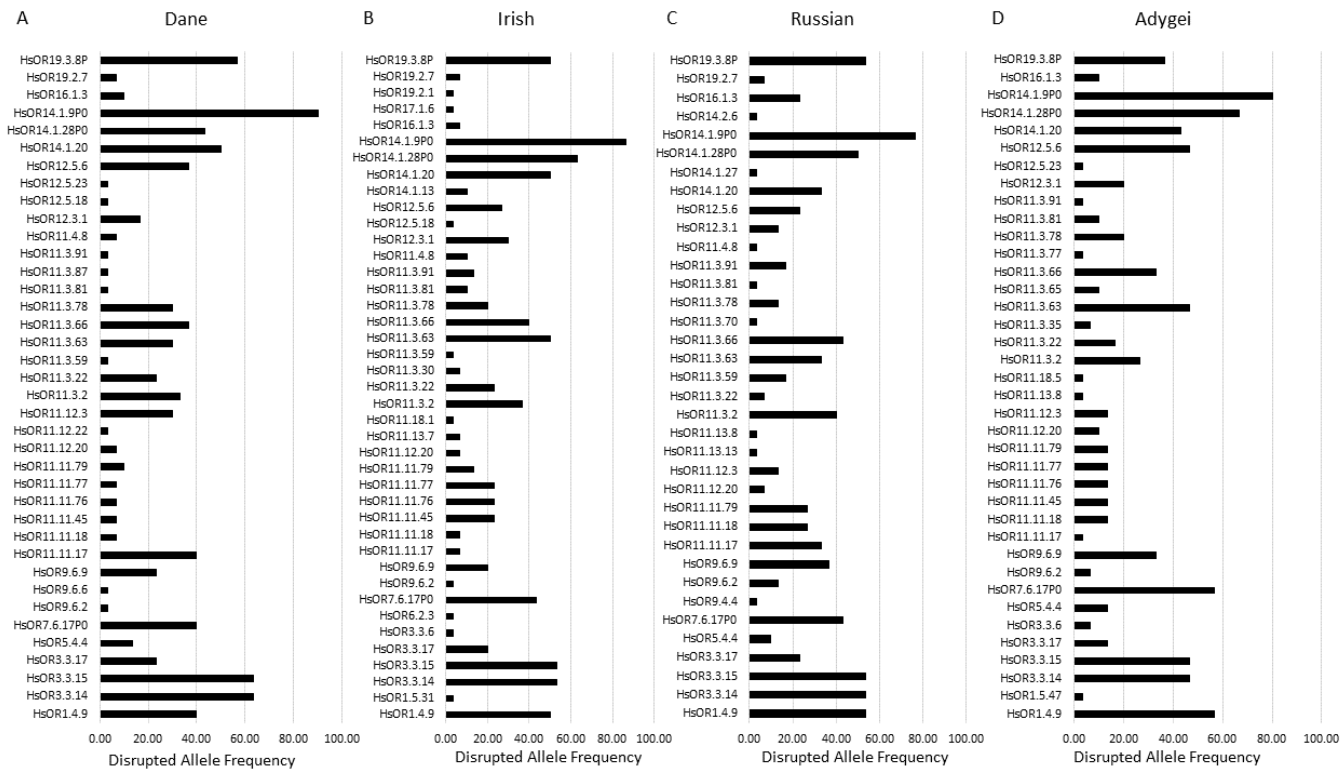

Figure 15: Disrupted Allele frequencies across study individuals of four European study populations. Given allele frequencies are in percentage and ordered in chromosomal order from bottom to top. Vertical axis represents gene name label while horizontal axis represents incidence of disrupted allele frequencies.

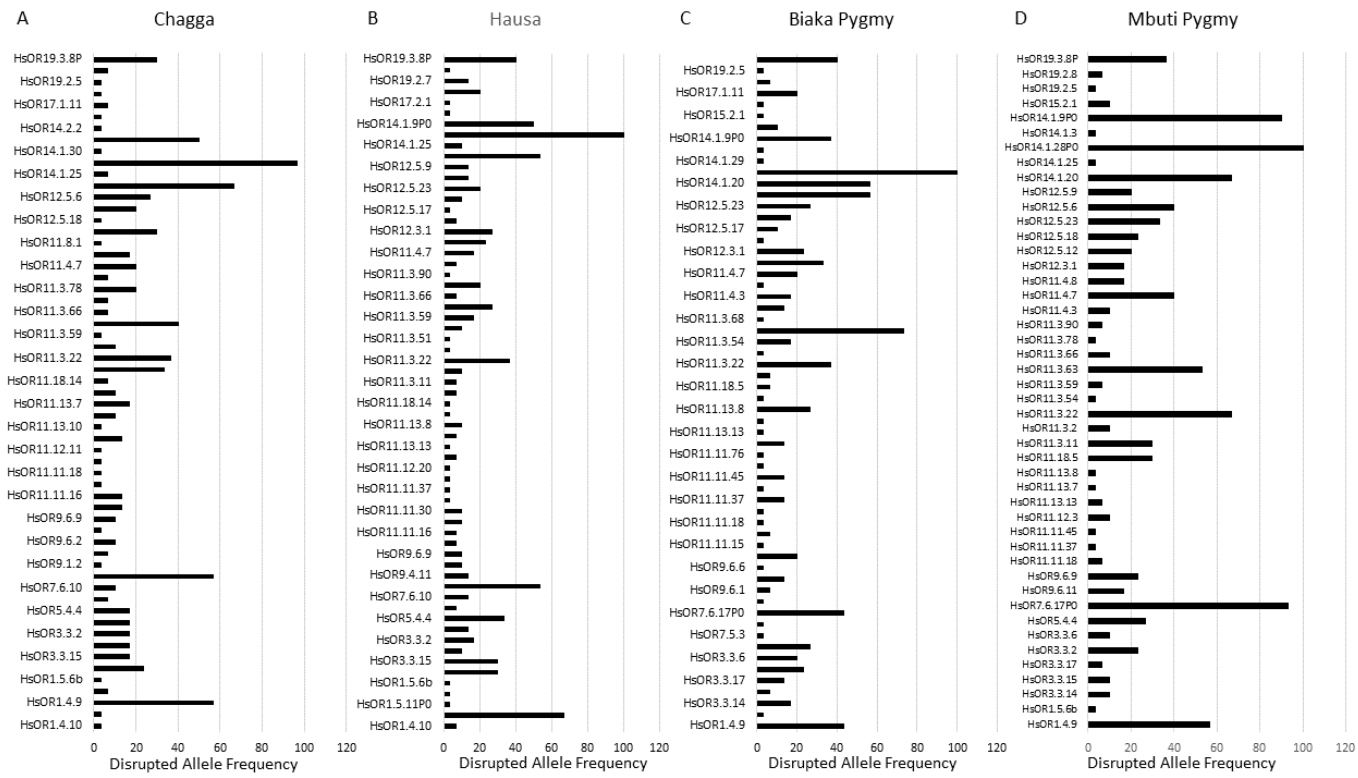

Figure 16: Disrupted Allele frequencies across study individuals of four African study populations. Given allele frequencies are in percentage and ordered in chromosomal order from bottom to top. Vertical axis represents gene name label while horizontal axis represents incidence of disrupted allele frequencies.

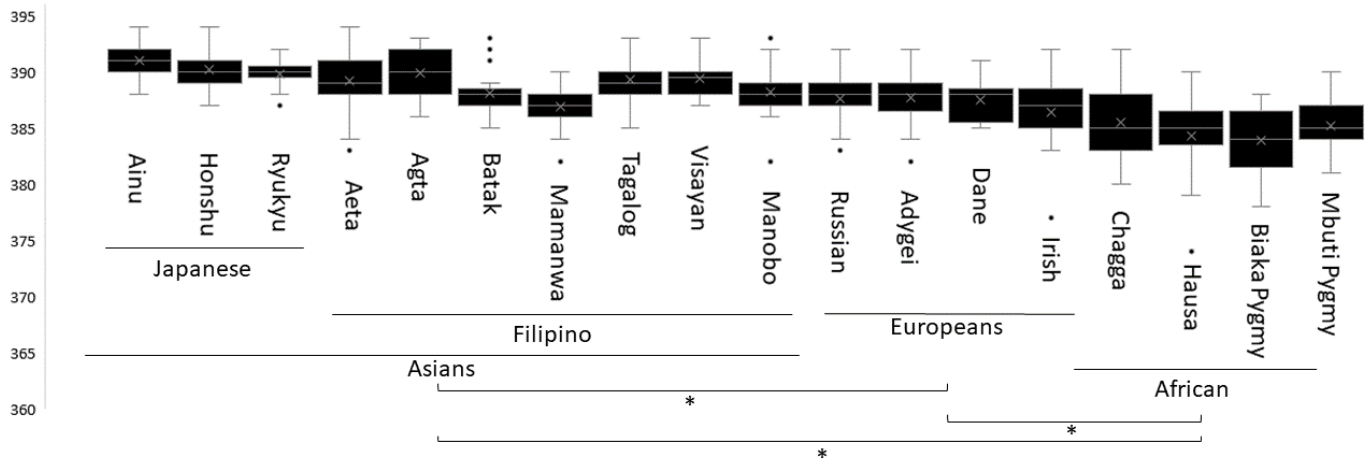

Figure 17: Distribution of number of homozygous intact OR genes among each study population. Number of homozygous intact OR genes in Asian populations are higher than European populations followed by African populations. * indicates a statistical significance.

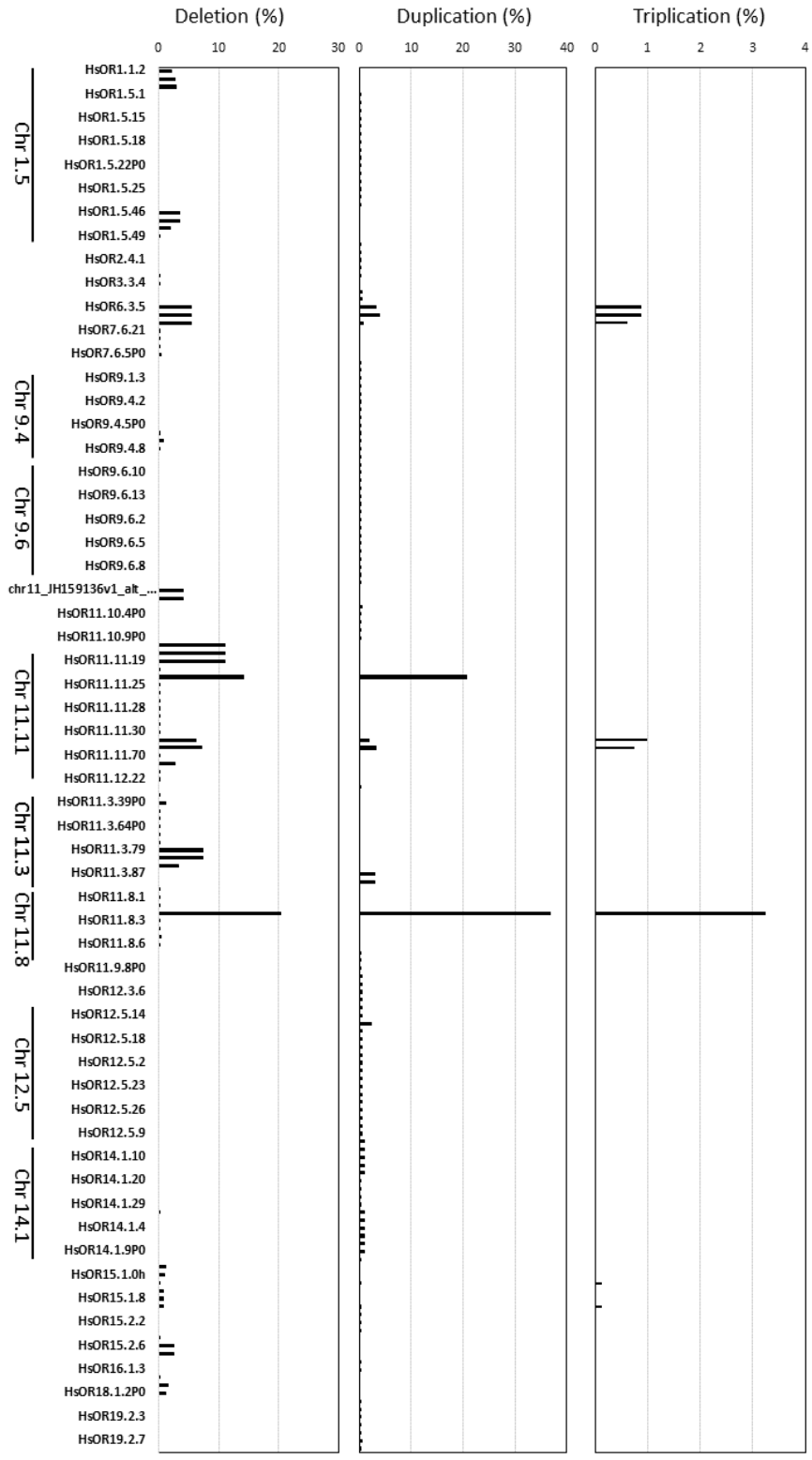

Figure 18: The deletion, duplication and triplication allele frequencies of autosomal OR genes. Due to limitation of space, all gene name labels are not added, however, CNV allele frequencies by gene are given in Appendix VII.

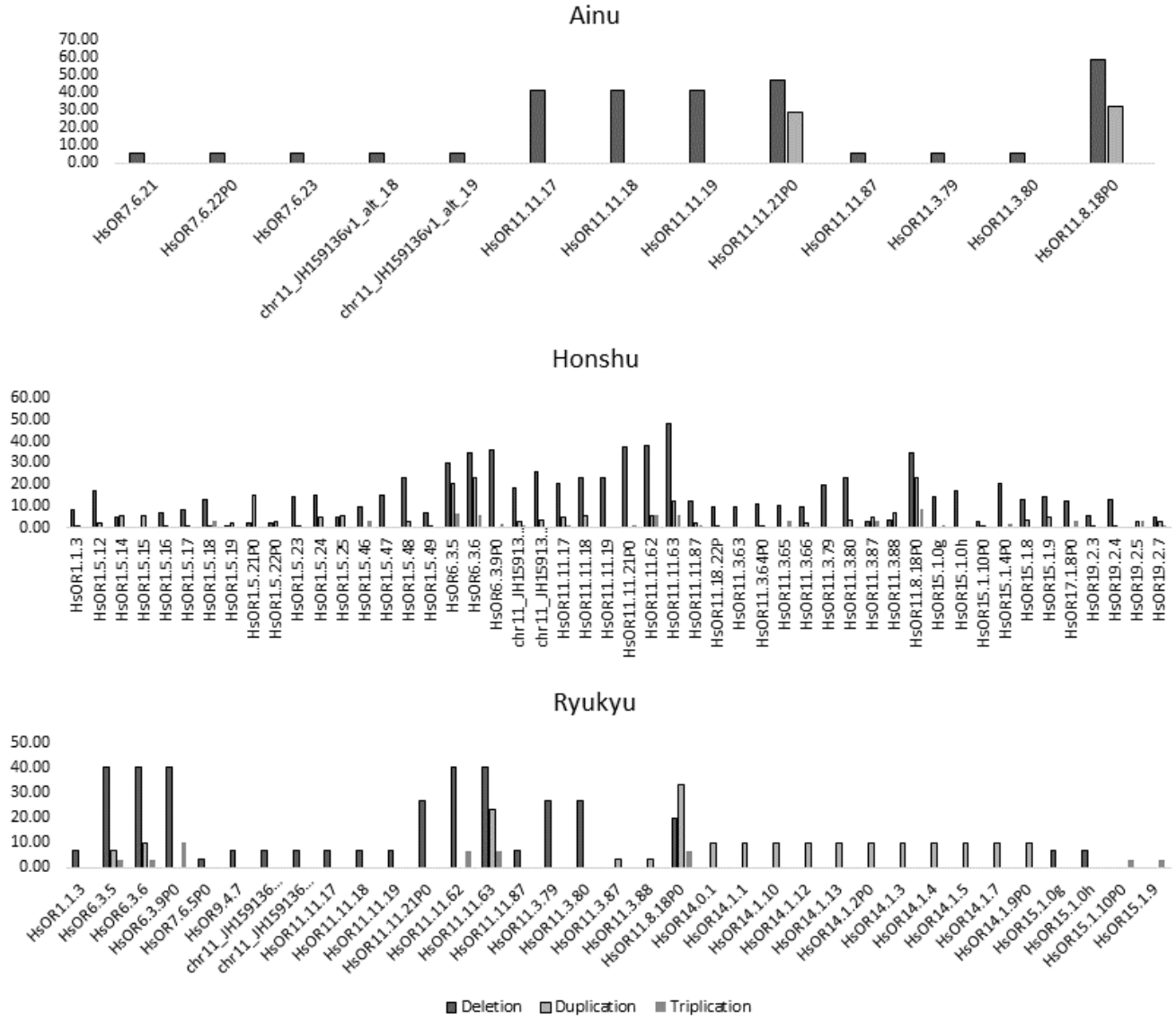

Figure 19: CNV (deletion, duplication, and triplication) allele frequencies in three Japanese study populations. Vertical axis represents incidence of color coded CNV allele in percentage while horizontal axis represents OR gene labels.

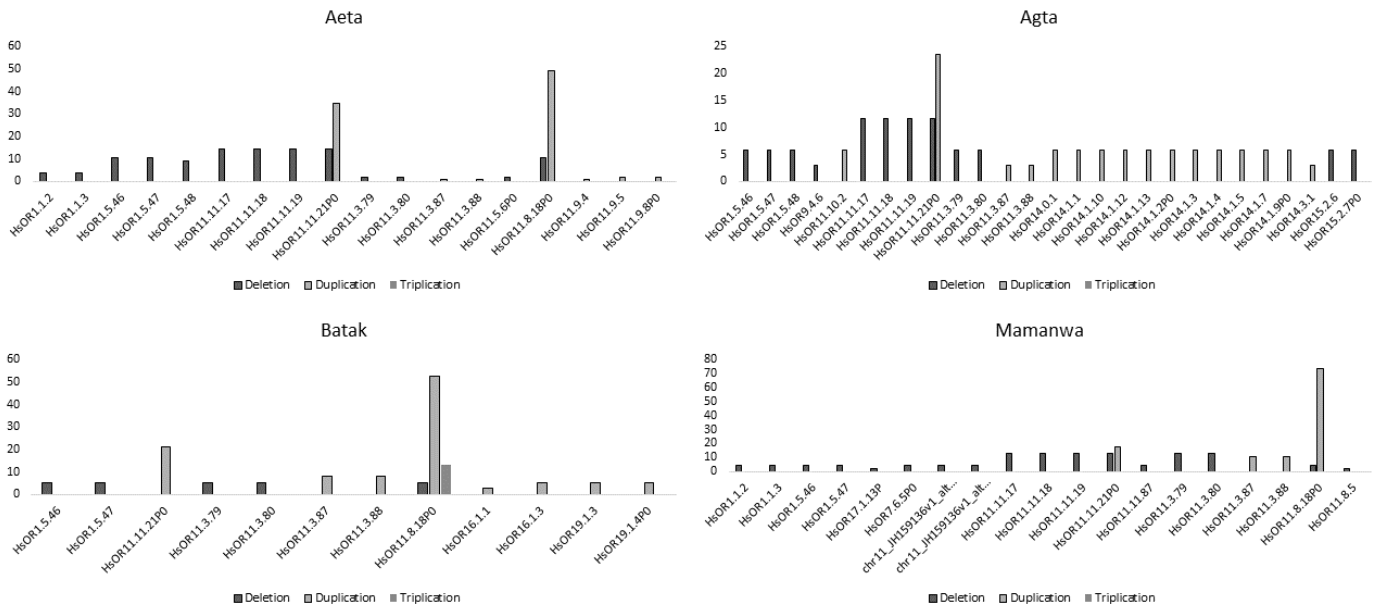

Figure 20: CNV (deletion, duplication, and triplication) allele frequencies in four negrito study populations of Philippines. Vertical axis represents incidence of color coded CNV allele in percentage while horizontal axis represents OR gene labels.

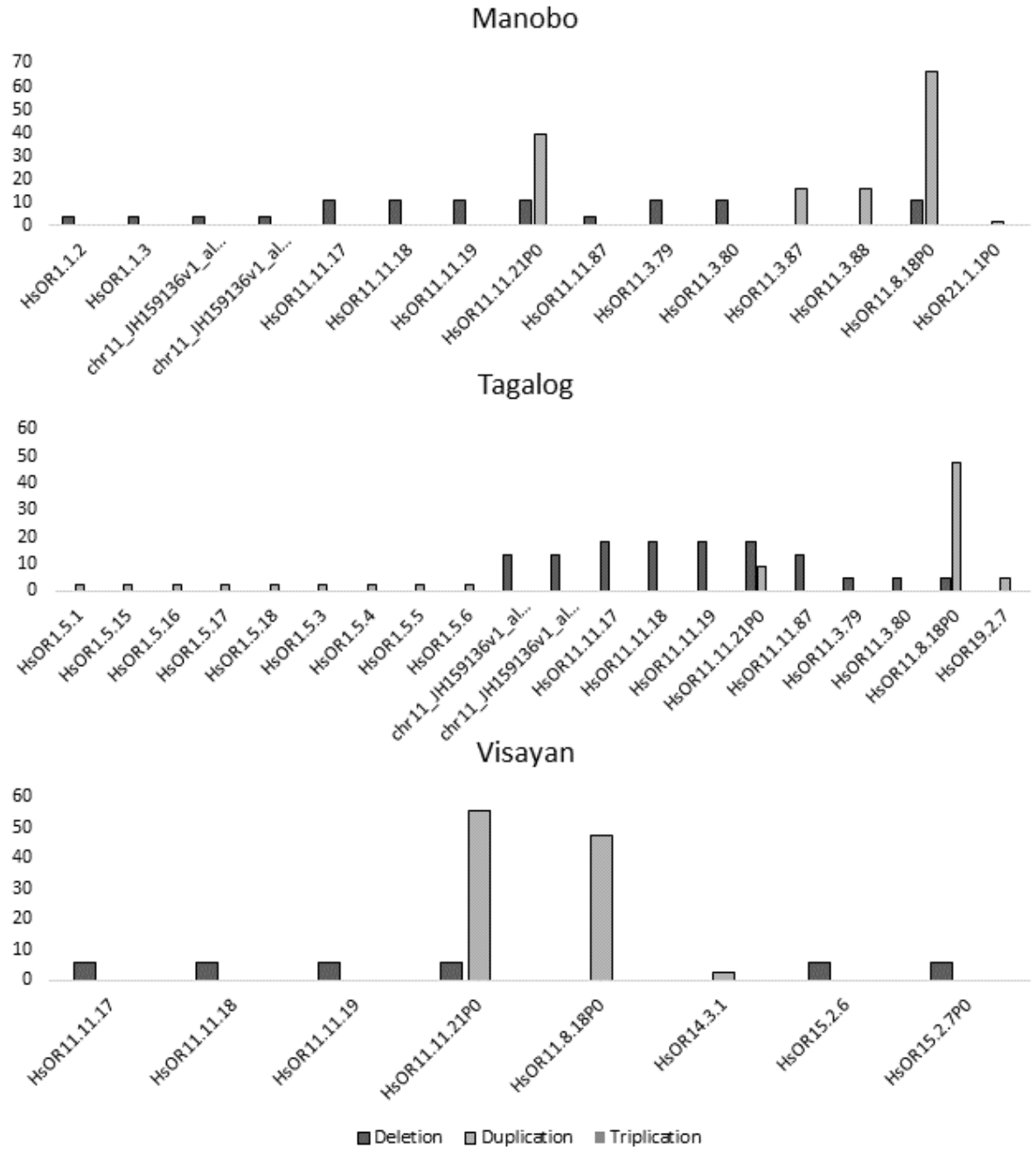

Figure 21: CNV (deletion, duplication, and triplication) allele frequencies in three agricultural non-negrito study populations of Philippines. Vertical axis represents incidence of color coded CNV allele in percentage while horizontal axis represents OR gene labels.

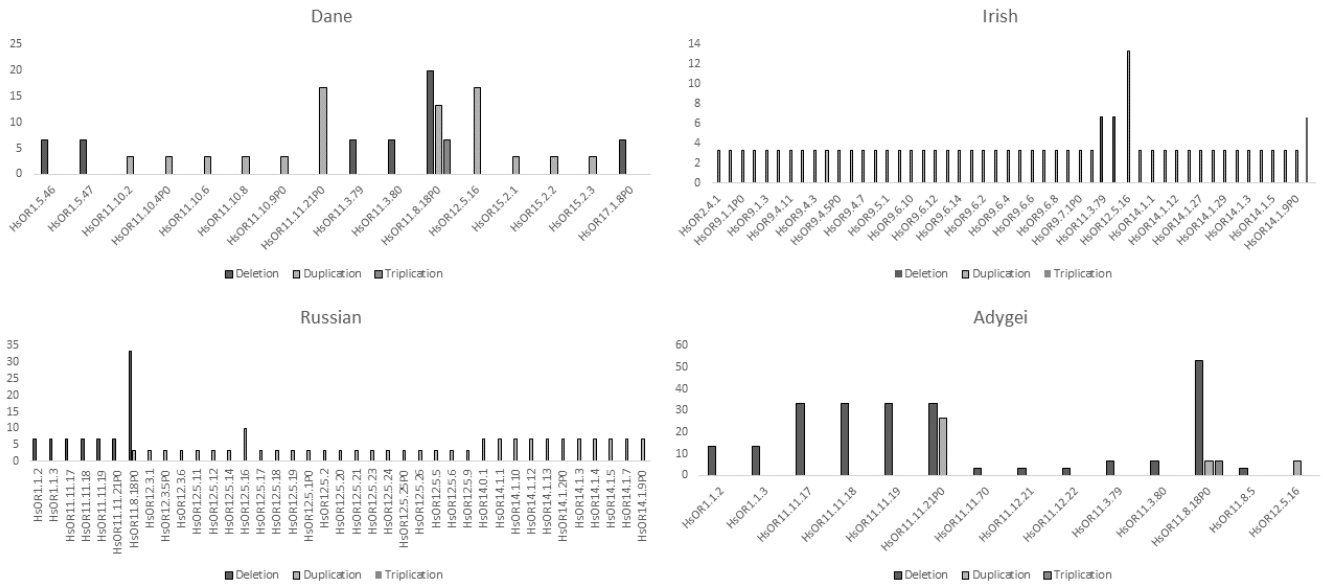

Figure 22: CNV (deletion, duplication, and triplication) allele frequencies in four European study populations. Vertical axis represents incidence of color coded CNV allele in percentage while horizontal axis represents OR gene labels

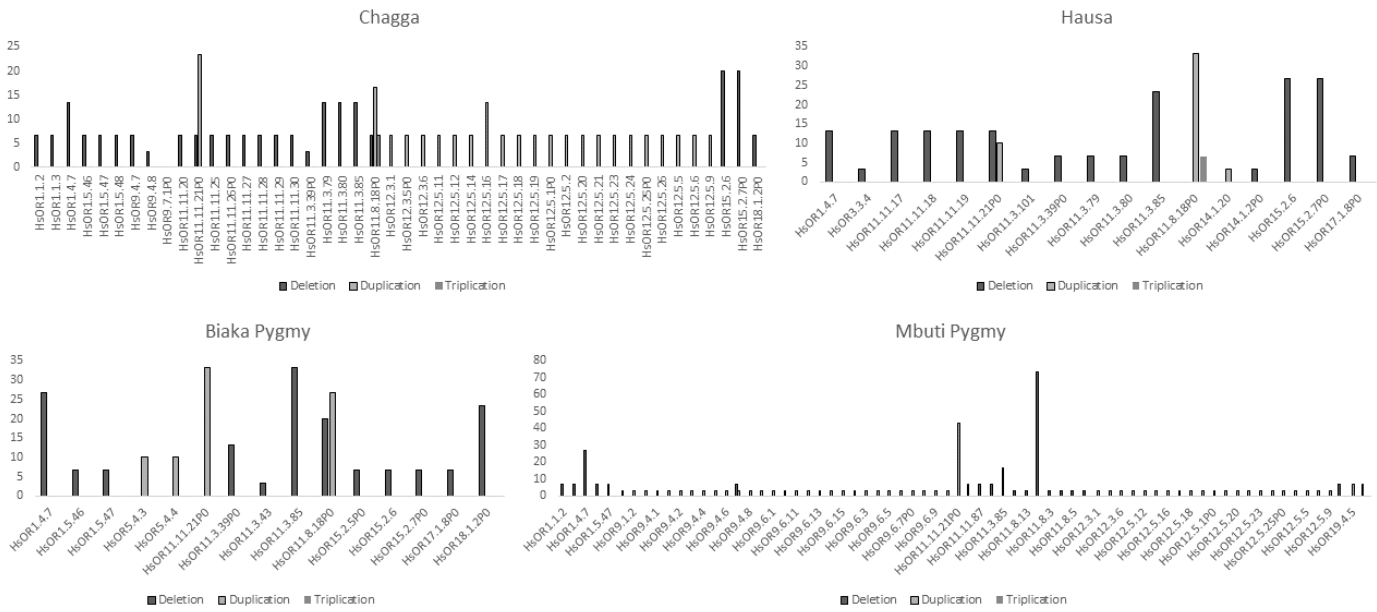

Figure 23: CNV (deletion, duplication, and triplication) allele frequencies in four African study populations. Vertical axis represents incidence of color coded CNV allele in percentage while horizontal axis represents OR gene labels

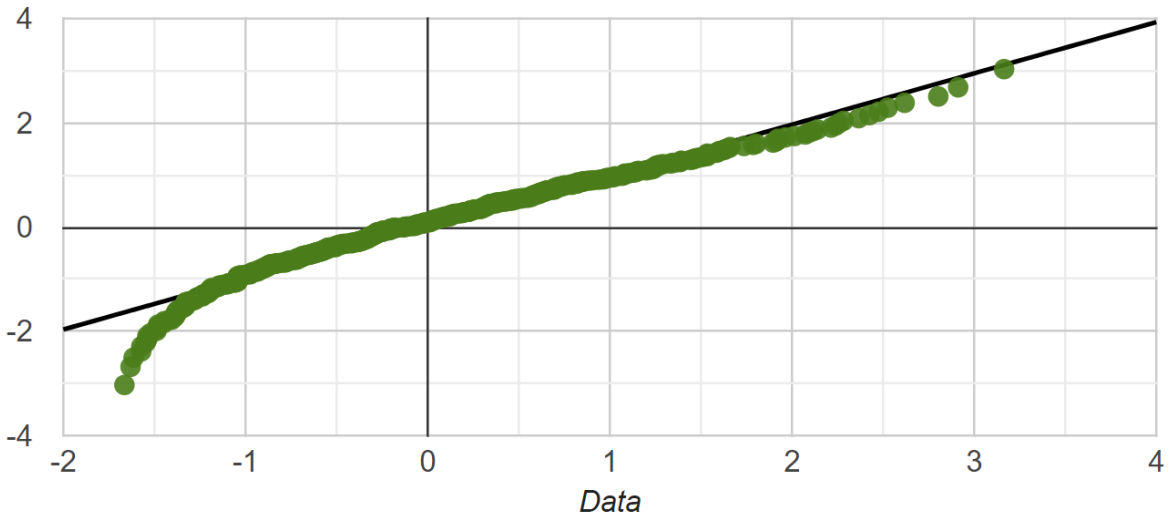

Figure 24: QQ plot to see distribution of Tajima's D among ORs and neutral references. Data can be seen to be in normal distribution although slightly deviated.

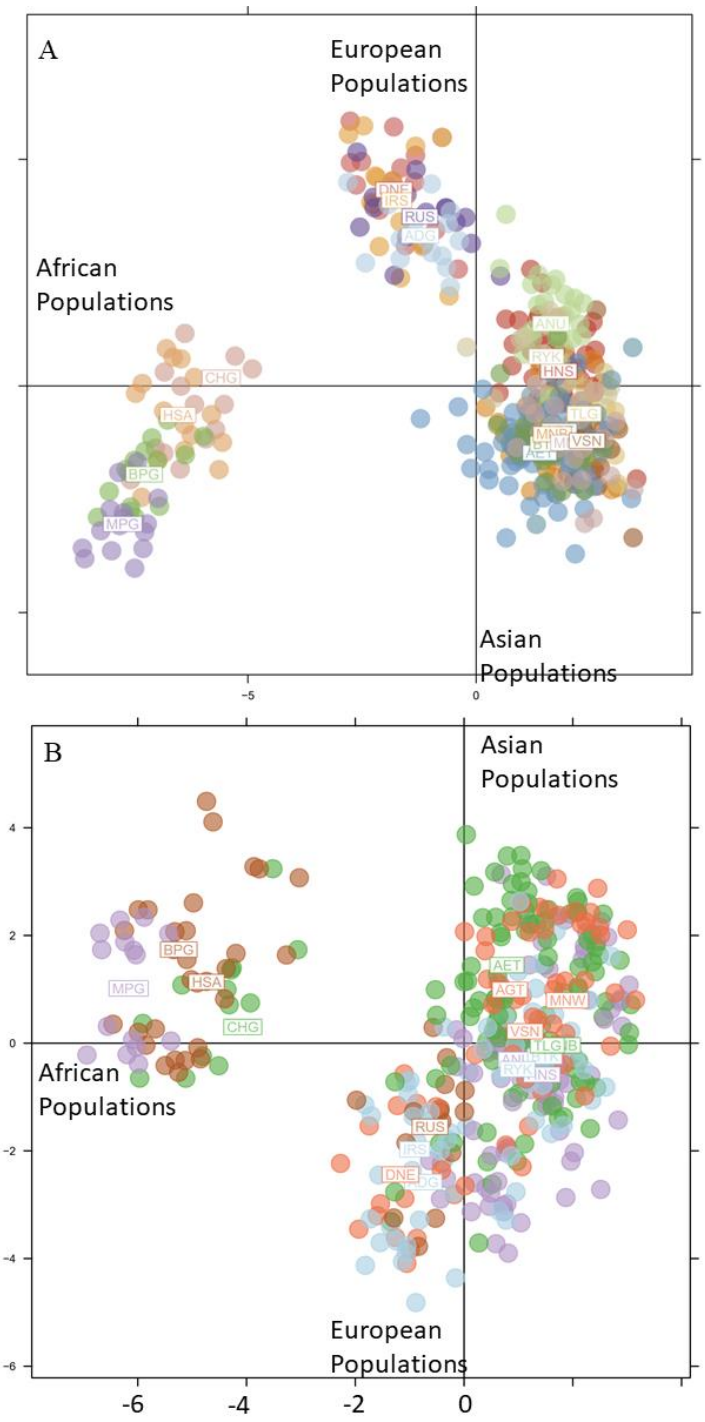

Figure 25: Population differentiation of ORs and neutral references. A: Population structure of ORs determined by SNP repertoire; B: Population structure of neutral references. Population codes in this analysis are BPG: Biaka Pygmy; MPG: Mbuti Pygmy; CHG: Chagga; HAS: Hausa; DNE: Dane; IRS: Irish; RUS: Russian; ADG: Adygei; ANU: Ainu; HNS: Honshu; RYK: Ryukyu; AET: Aeta; AGT: Agta; BTK: Batak; MNW: Mamanwa; MNB: Manobo; TLG: Tagalog; VSN: Visayan

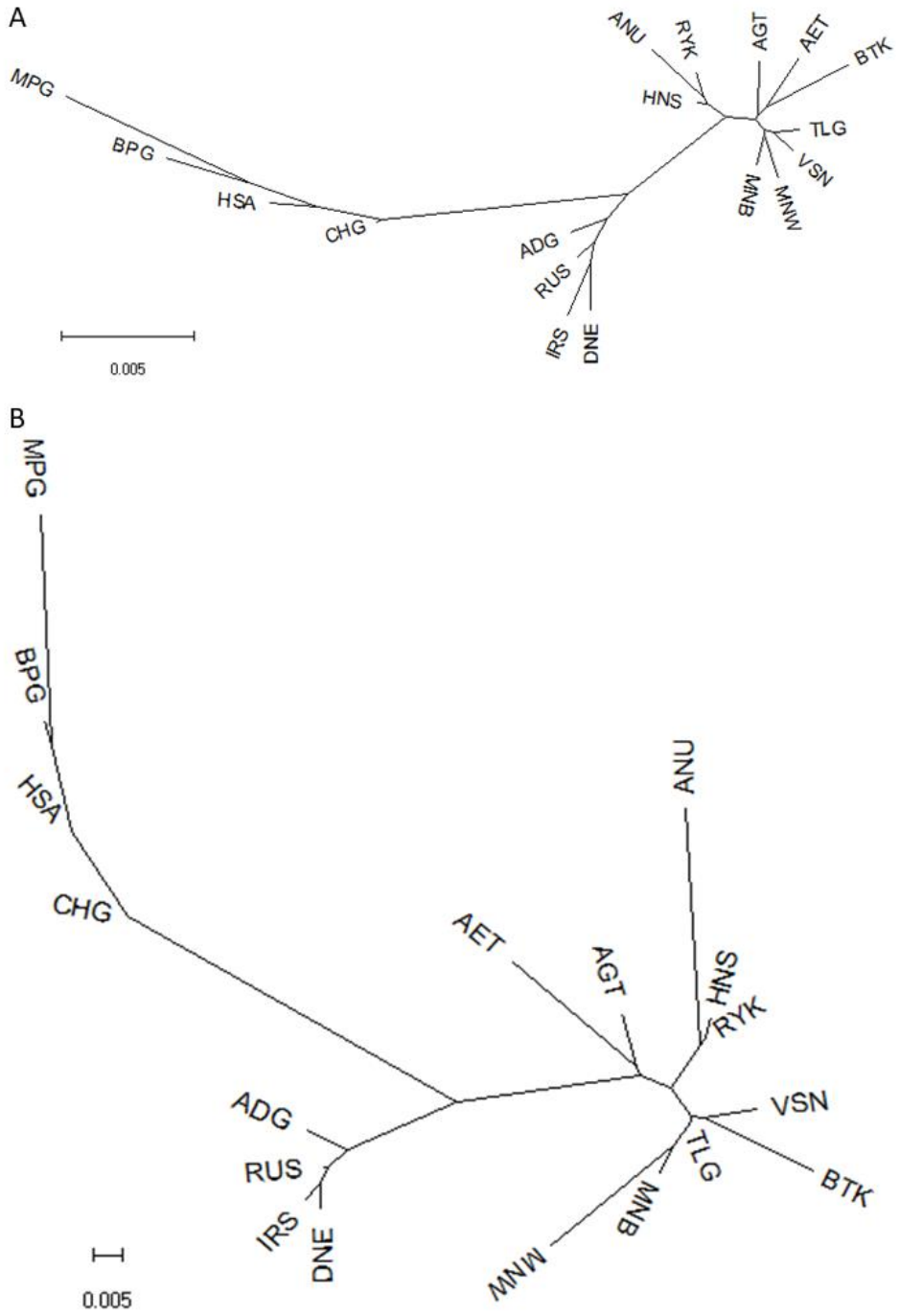

Figure 26: N_{ST} genetic differentiation as a quantitative measure of population differentiation. A: ORs; B: neutral reference. Population codes in this figrue are BPG: Biaka Pygmy; MPG: Mbuti Pygmy; CHG: Chagga; HAS: Hausa; DNE: Dane; IRS: Irish; RUS: Russian; ADG: Adygei; ANU: Ainu; HNS: Honshu; RYK: Ryukyu; AET: Aeta; AGT: Agta; BTK: Batak; MNW: Mamanwa; MNB: Manobo; TLG: Tagalog; VSN: Visayan

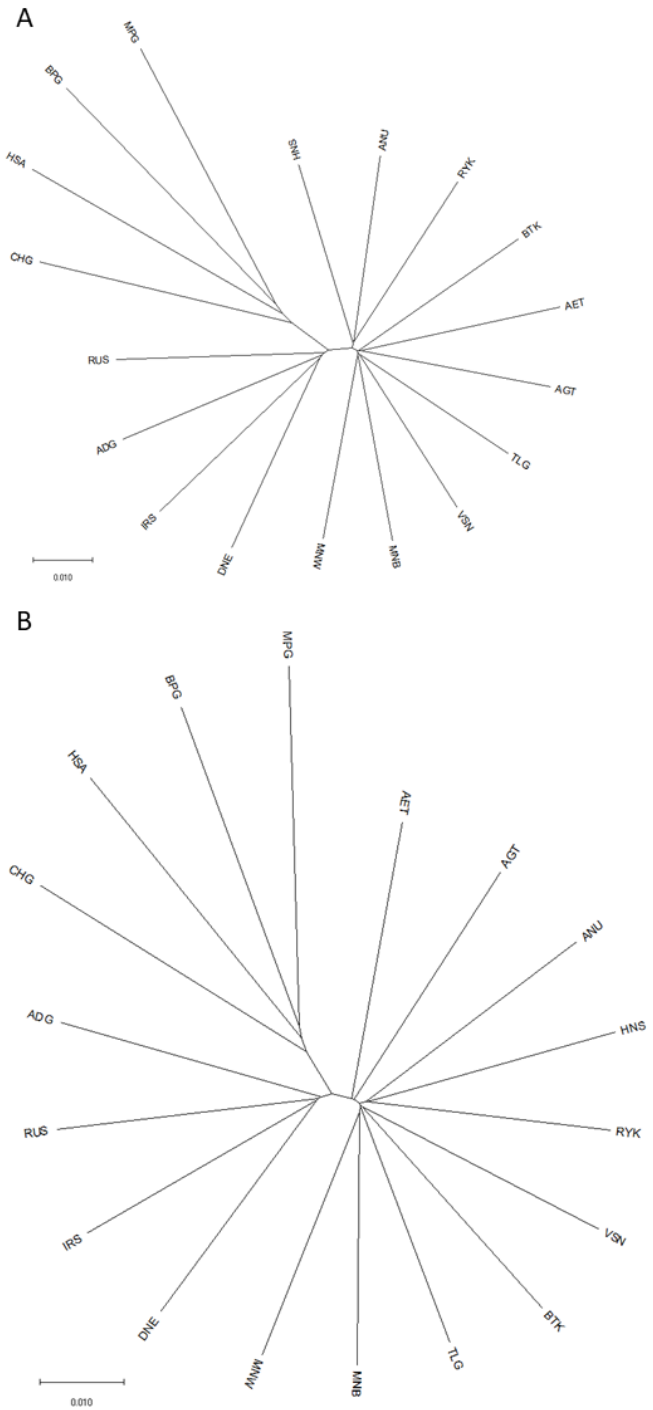

Figure 27: d_{xy} distance matrix as a quantitative measure of population differentiation. A: ORs; B: neutral reference. Population codes in this figrue are BPG: Biaka Pygmy; MPG: Mbuti Pygmy; CHG: Chagga; HAS: Hausa; DNE: Dane; IRS: Irish; RUS: Russian; ADG: Adygei; ANU: Ainu; HNS: Honshu; RYK: Ryukyu; AET: Aeta; AGT: Agta; BTK: Batak; MNW: Mamanwa; MNB: Manobo; TLG: Tagalog; VSN: Visayan

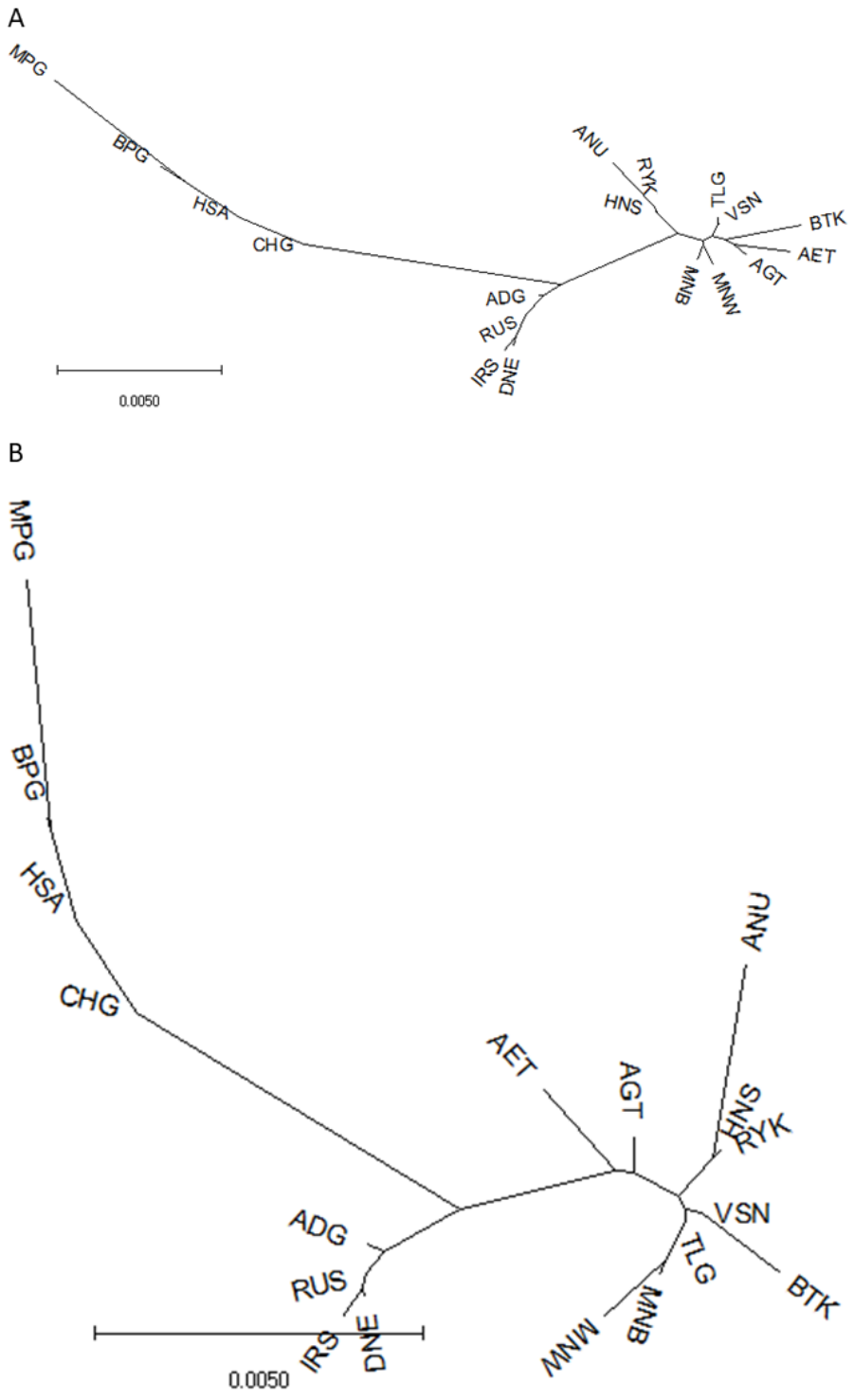

Figure 28: d_A distance matrix as a quantitative measure of population differentiation. A: ORs; B: neutral reference. Population codes in this figrue are BPG: Biaka Pygmy; MPG: Mbuti Pygmy; CHG: Chagga; HAS: Hausa; DNE: Dane; IRS: Irish; RUS: Russian; ADG: Adygei; ANU: Ainu; HNS: Honshu; RYK: Ryukyu; AET: Aeta; AGT: Agta; BTK: Batak; MNW: Mamanwa; MNB: Manobo; TLG: Tagalog; VSN: Visayan

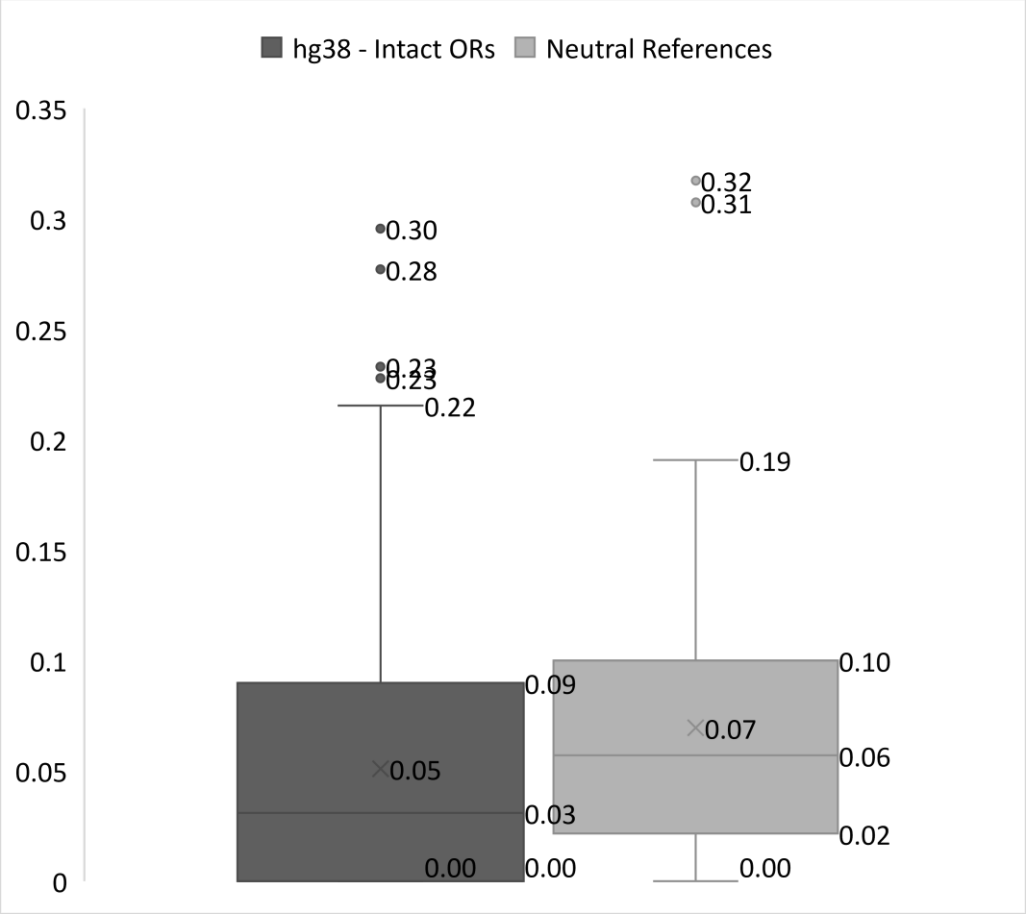

Figure 29: Distribution of N_{ST} values per gene/region among autosomal hg38 – intact ORs and neutral references.

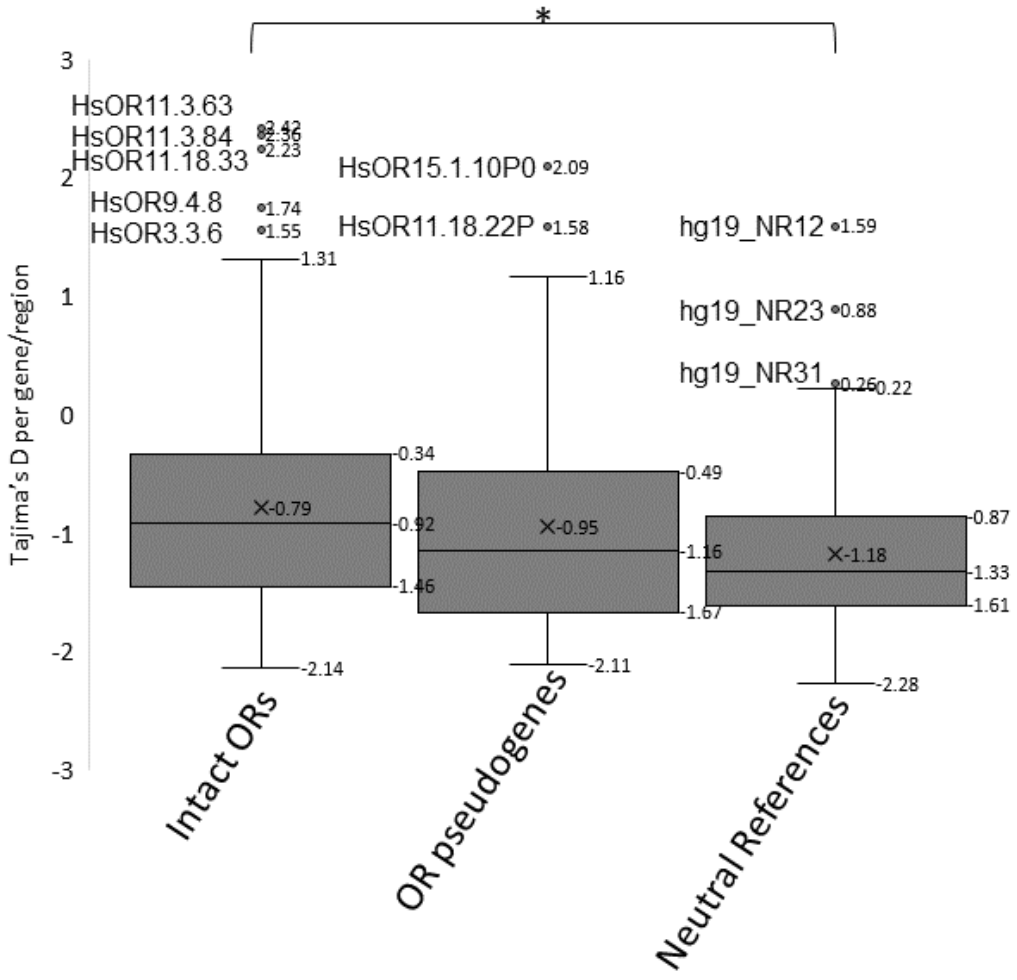

Figure 30: Tajima's D of hg38 OR intact and pseudogenes and neutral references examined for all 401 study individuals. Outlier OR genes and neutral references are labelled. * indicates statistical significant difference.

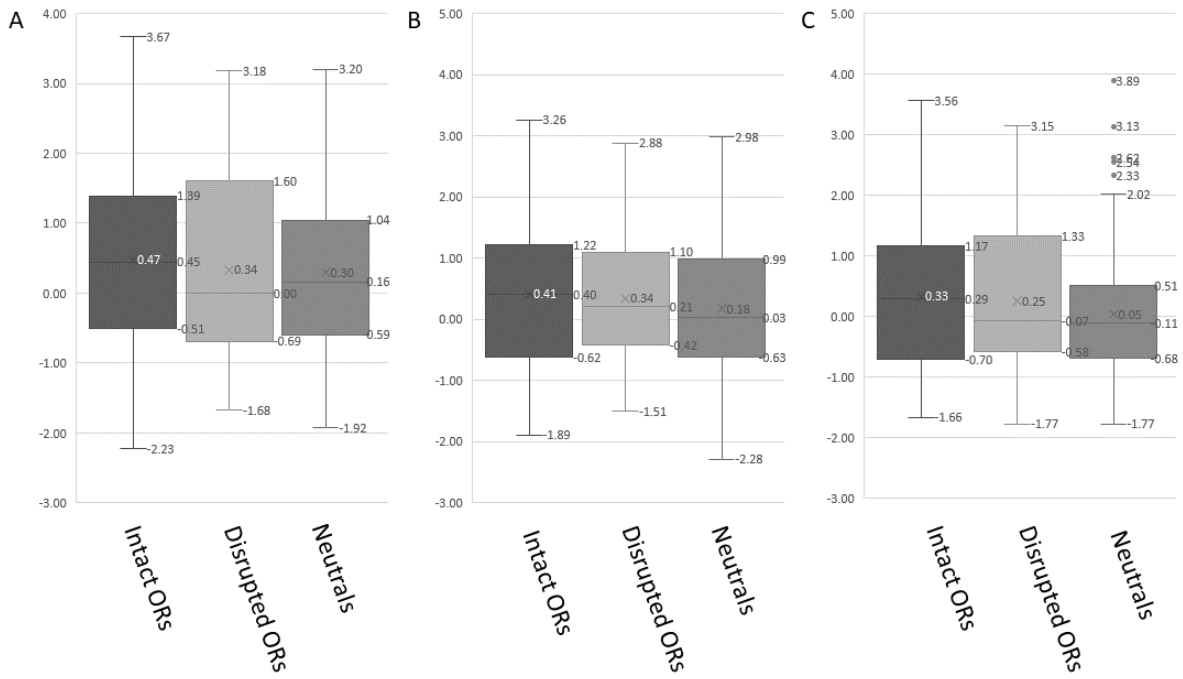

Figure 31: Tajima's D of intact ORs, disrupted ORs and neutral references. A: Honshu study individuals; B: Ryukyu study individuals; C: Ainu study individuals

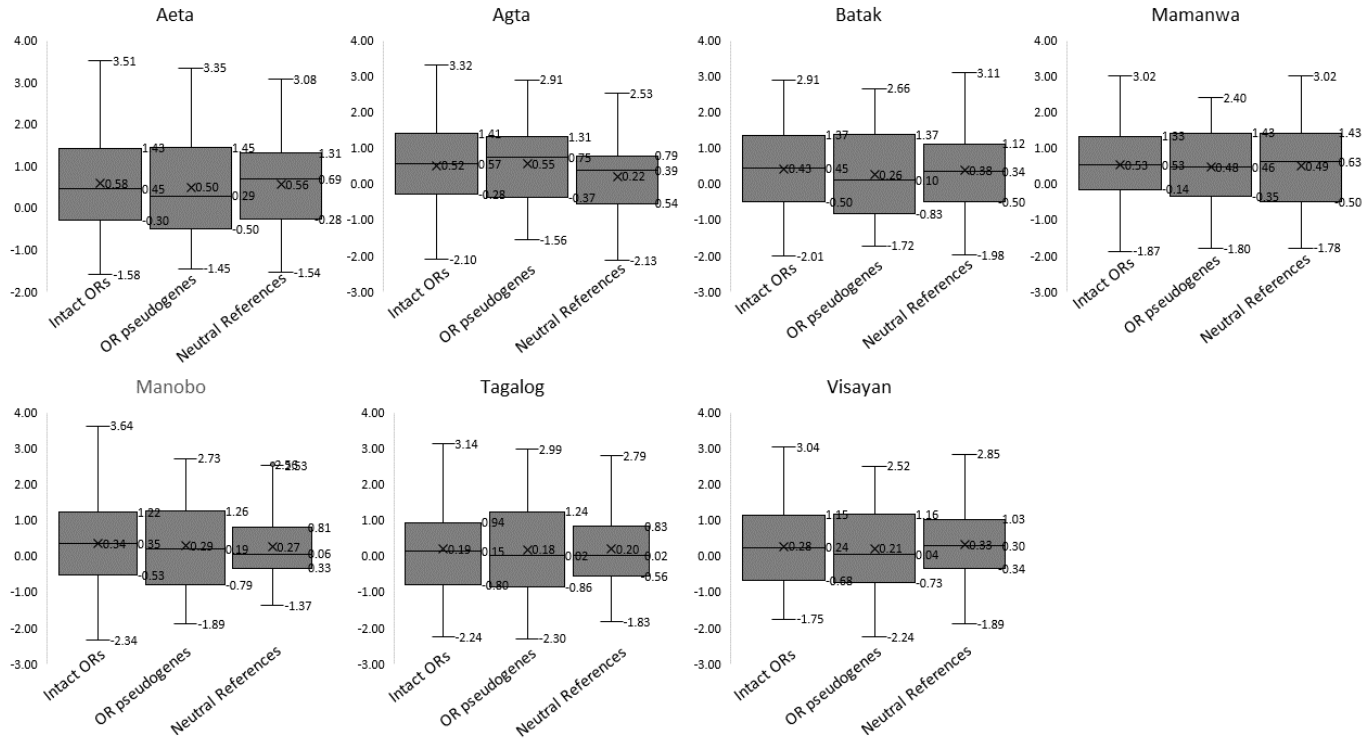

Figure 32: Tajima's *D* distribution of OR intact, pseudogenes and neutral references in Filipino study populations. First row represents four negrito populations while the second row represents non-negrito agriculturists of Philippines. Tajima's D per gene/neutral reference is used to make boxplots.

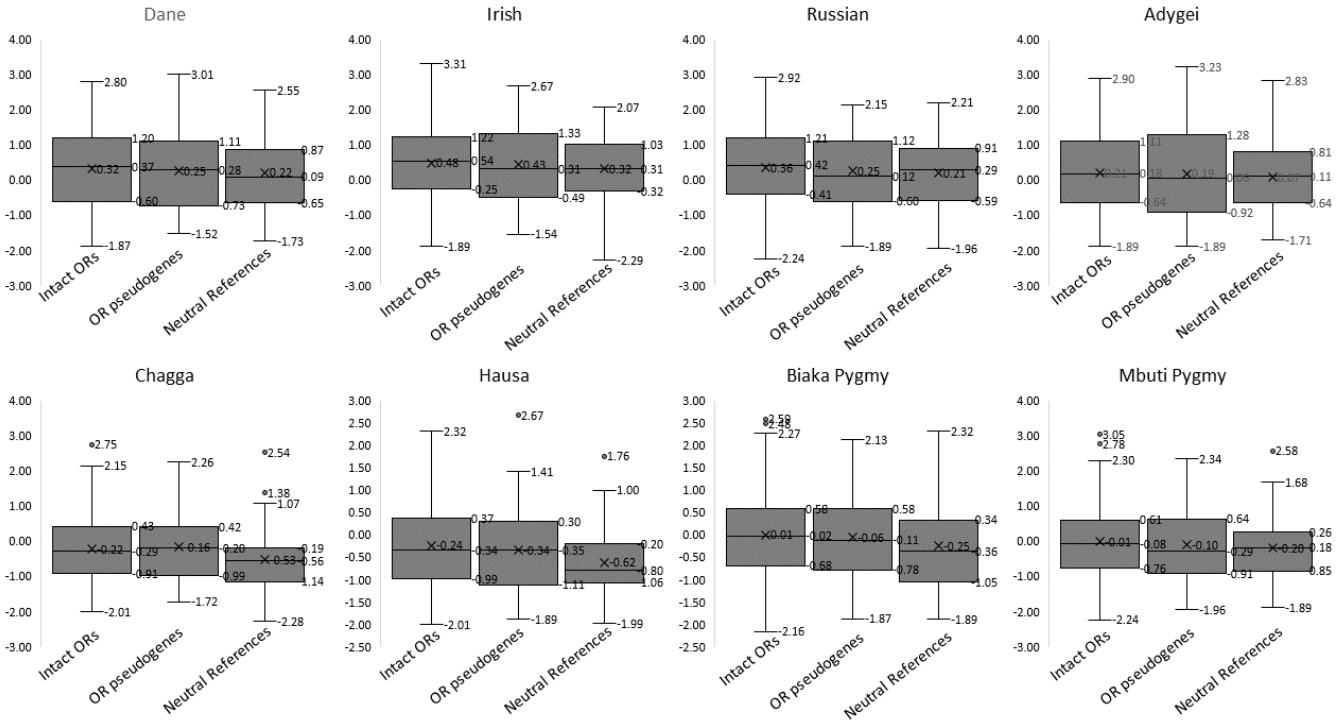

Figure 33: Tajima's D distribution of OR intact, pseudogenes and neutral references in European and African study populations. First row represents four European populations while the second row represents four African populations. Tajima's D per gene/neutral reference is used to make boxplots.

REFERENCES

Adachi, N., Kanzawa-Kiriyama, H., Nara, T., Kakuda, T., Nishida, I. & Shinoda, K.-I. 2021. Ancient genomes from the initial Jomon period: new insights into the genetic history of the Japanese archipelago. *Anthropological Science,* 129**,** 13-22.

Adipietro, K. A., Mainland, J. D. & Matsunami, H. 2012. Functional evolution of Mammalian odorant receptors. *PLoS Genetics,* 8**,** e1002821.

Akhtar, M. S., Ashino, R., Oota, H., Ishida, H., Niimura, Y., Touhara, K., Melin, A. D. & Kawamura, S. 2022. Genetic variation of olfactory receptor gene family in a Japanese population. *Anthropological Science***,** 211024.

Amadori, E., Scala, M., Cereda, G. S., Vari, M. S., Marchese, F., Di Pisa, V., Mancardi, M. M., Giacomini, T., Siri, L. & Vercellino, F. 2020. Targeted re-sequencing for early diagnosis of genetic causes of childhood epilepsy: the Italian experience from the 'beyond epilepsy'project. *Italian journal of pediatrics,* 46**,** 1-9.

Arbiza, L., Zhong, E. & Keinan, A. 2012. NRE: a tool for exploring neutral loci in the human genome. *BMC Bioinformatics,* 13**,** 301.

Bannai, M., Ohashi, J., Harihara, S., Takahashi, Y., Juji, T., Omoto, K. & Tokunaga, K. 2000. Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens,* 55**,** 128-139.

Bannai, M., Tokunaga, K., Imanishi, T., Harihara, S., Fujisawa, K., Juji, T. & Omoto, K. 1996. HLA class II alleles in Ainu living in Hidaka District, Hokkaido, northern Japan. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists,* 101**,** 1-9.

Braun, T., Voland, P., Kunz, L., Prinz, C. & Gratzl, M. 2007. Enterochromaffin cells of the human gut: sensors for spices and odorants. *Gastroenterology,* 132**,** 1890-1901.

Buck, L. & Axel, R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell,* 65**,** 175-87.

Bushdid, C., Magnasco, M. O., Vosshall, L. B. & Keller, A. 2014. Humans can discriminate more than 1 trillion olfactory stimuli. *Science,* 343**,** 1370-2.

Carithers, L. J. & Moore, H. M. 2015. The genotype-tissue expression (GTEx) project. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

Cavalli-Sforza, L. 1986. African Pygmies: an evaluation of the state of research. *African pygmies***,** 361-426.

Chang, A. J., Ortega, F. E., Riegler, J., Madison, D. V. & Krasnow, M. A. 2015. Oxygen regulation of breathing through an olfactory receptor activated by lactate. *Nature,* 527**,** 240-244.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly,* 6**,** 80-92.

Concas, M. P., Cocca, M., Francescatto, M., Battistuzzi, T., Spedicati, B., Feresin, A., Morgan, A., Gasparini, P. & Girotto, G. 2021. The role of knockout olfactory receptor genes in odor discrimination. *Genes,* 12**,** 631.

Consortium, T. G. P. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

Consortium, T. G. P. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

Consortium, T. I. H. 2005. A haplotype map of the human genome. *Nature,* 437**,** 1299-320.

Consortium, T. I. H. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449**,** 851-61.

Cook, D. E. & Andersen, E. C. 2017. VCF-kit: assorted utilities for the variant call format. *Bioinformatics,* 33**,** 1581-1582.

Coordinators, N. R. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research,* 44**,** D7-19.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A. & Davies, R. M. 2021. Twelve years of SAMtools and BCFtools. *Gigascience,* 10**,** giab008.

Dávila, N., Shea, B. T., Omoto, K., Mercado, M., Misawa, S. & Baumann, G. 2002. Growth hormone binding protein, insulin-like growth factor-I and short stature in two pygmy populations from the Philippines. *Journal of Pediatric Endocrinology and Metabolism,* 15**,** 269-276.

DeMaria, S. & Ngai, J. 2010. The cell biology of smell. *Journal of Cell Biology,* 191**,** 443-452.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics,* 43**,** 491-8.

Eaton, K. M., Bernal, M. A., Backenstose, N. J. C., Yule, D. L. & Krabbenhoft, T. J. 2021. Nanopore Amplicon Sequencing Reveals Molecular Convergence and Local Adaptation of Rhodopsin in Great Lakes Salmonids. *Genome Biology and Evolution,* 13**,** evaa237.

Fujita, Y., Takahashi, T., Suzuki, A., Kawashima, K., Nara, F. & Koishi, R. 2007. Deorphanization of Dresden G protein-coupled receptor for an odorant receptor. *Journal of Receptors and Signal Transduction,* 27**,** 323-334.

Gakuhari, T., Nakagome, S., Rasmussen, S., Allentoft, M. E., Sato, T., Korneliussen, T., Chuinneagain, B. N., Matsumae, H., Koganebuchi, K., Schmidt, R., Mizushima, S., Kondo, O., Shigehara, N., Yoneda, M., Kimura, R., Ishida, H., Masuyama, T., Yamada, Y., Tajima, A., Shibata, H., Toyoda, A., Tsurumoto, T., Wakebe, T., Shitara, H., Hanihara, T., Willerslev, E., Sikora, M. & Oota, H. 2020. Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. *Communications Biology,* 3**,** 437.

Geldard, F. A., O'Hehir, R. & Gavens, D. 1953. *The human senses,* Wiley New York.

Gil-Varea, E., Spataro, N., Villar, L. M., Tejeda-Velarde, A., Midaglia, L., Matesanz, F., Malhotra, S., Eixarch, H., Patsopoulos, N. & Fernández, Ó. 2020. Targeted resequencing reveals rare variants enrichment in multiple sclerosis susceptibility genes. *Human mutation,* 41**,** 1308-1320.

Gilad, Y. & Lancet, D. 2003. Population differences in the human functional olfactory repertoire. *Molecular Biology and Evolution,* 20**,** 307-314.

Glusman, G., Yanai, I., Rubin, I. & Lancet, D. 2001. The complete human olfactory subgenome. *Genome Research,* 11**,** 685-702.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. & Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology,* 27**,** 182-9.

Go, Y. & Niimura, Y. 2008. Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. *Molecular biology and evolution,* 25**,** 1897-1907.

Greenberg, J. H. 1947. Islam and clan organization among the Hausa. *Southwestern journal of Anthropology,* 3**,** 193-211.

Grossi, A., Severino, M., Rusmini, M., Tortora, D., Ramenghi, L. A., Cama, A., Rossi, A., Di Rocco, M., Ceccherini, I. & Bertamino, M. 2020. Targeted re-sequencing in pediatric and perinatal stroke. *European Journal of Medical Genetics,* 63**,** 104030.

Halperin, C. J. 1983. Russia in the Mongol Empire in Comparative Perspective. *Harvard Journal of Asiatic Studies,* 43**,** 239-261.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. & Horai, S. 2006. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *Journal of human genetics,* 51**,** 47-58.

Hanihara, K. 1991. Dual structure model for the population history of the Japanese. *Japan Review,* 2**,** 1-33.

Harihara, S., Hirai, M. & Omoto, K. 1986. Mitochondrial DNA polymorphism in Japanese living in Hokkaido. *Japanese journal of human genetics,* 31**,** 73-83.

Harihara, S., Saitou, N., Hirai, M., Gojobori, T., Park, K., Misawa, S., Ellepola, S., Ishida, T. & Omoto, K. 1988. Mitochondrial DNA polymorphism among five Asian populations. *American journal of human genetics,* 43**,** 134.

Hasin, Y., Olender, T., Khen, M., Gonzaga-Jauregui, C., Kim, P. M., Urban, A. E., Snyder, M., Gerstein, M. B., Lancet, D. & Korbel, J. O. 2008. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genetics,* 4**,** e1000249.

Heymer, A. 1980. BAYAKA-PYGMIES (CENTRAL-AFRICA)-GIVING, TAKING, SHARING AND RANK ORDER DEMONSTRATION IN RELATION WITH THESE BEHAVIOR PATTERNS. *HOMO,* 31**,** 252-265.

Hiramatsu, C., Melin, A. D., Aureli, F., Schaffner, C. M., Vorobyev, M. & Kawamura, S. 2009. Interplay of olfaction and vision in fruit foraging of spider monkeys. *Animal Behaviour,* 77**,** 1421-1426.

Hoover, K. C. 2010. Smell with inspiration: the evolutionary significance of olfaction. *Yearbook of Physical Anthropology,* 53**,** 63–74.

Hoover, K. C., Gokcumen, O., Qureshy, Z., Bruguera, E., Savangsuksa, A., Cobb, M. & Matsunami, H. 2015. Global survey of variation in a human olfactory receptor gene reveals signatures of non-neutral evolution. *Chemical Senses,* 40**,** 481-8.

Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K. S., Omoto, K. & Pan, I.-H. 1996. mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *American journal of human genetics,* 59**,** 579.

Horai, S., Omoto, K., Juji, T., Sonozaki, H., Mitsui, H., Misawa, S., Sumpaico, J. S. & Mercado, A. S. 1981. The HLA antigens of two Negrito populations in the Philippines. *Tissue Antigens,* 17**,** 343-348.

Hughes, G. M., Teeling, E. C. & Higgins, D. G. 2014. Loss of olfactory receptor function in hominin evolution. *PLoS ONE,* 9**,** e84714.

Iadarola, B., Xumerle, L., Lavezzari, D., Paterno, M., Marcolungo, L., Beltrami, C., Fortunati, E., Mei, D., Vetro, A. & Guerrini, R. 2020. Shedding light on dark genes: enhanced targeted resequencing by optimizing the combination of enrichment technology and DNA fragment length. *Scientific reports,* 10**,** 1-11.

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., Jahesh, G., Khan, H., Coombe, L., Warren, R. L. & Birol, I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Research,* 27**,** 768-777.

Jaeger, S. R., McRae, J. F., Bava, C. M., Beresford, M. K., Hunter, D., Jia, Y., Chheang, S. L., Jin, D., Peng, M. & Gamble, J. C. 2013a. A Mendelian trait for olfactory

sensitivity affects odor experience and food selection. *Current Biology,* 23**,** 1601-1605.

Jaeger, S. R., McRae, J. F., Bava, C. M., Beresford, M. K., Hunter, D., Jia, Y., Chheang, S. L., Jin, D., Peng, M., Gamble, J. C., Atkinson, K. R., Axten, L. G., Paisley, A. G., Tooman, L., Pineau, B., Rouse, S. A. & Newcomb, R. D. 2013b. A mendelian trait for olfactory sensitivity affects odor experience and food selection. *Current Biology,* 23**,** 1601-5.

Jensen, R. A. 2001. Orthologs and paralogs-we need to get it right. *Genome biology,* 2**,** 1-3.

Jimenez, R. C., Casajuana-Martin, N., García-Recio, A., Alcántara, L., Pardo, L., Campillo, M. & Gonzalez, A. 2021. The mutational landscape of human olfactory G protein-coupled receptors. *BMC biology,* 19**,** 1-15.

Jin, W., Lei, Z., Xu, S., Fachen, Z., Yixiang, Z., Shilei, Z., Tao, G., Zhe, S., Fengzhou, L. & Su, W.-H. 2021. Genetic Mutation Analysis in Small Cell Lung Cancer by a Novel NGS-Based Targeted Resequencing Gene Panel and Relation with Clinical Features. *BioMed research international,* 2021.

Jinam, T. A., Kanzawa-Kiriyama, H. & Saitou, N. 2015. Human genetic diversity in the Japanese Archipelago: dual structure and beyond. *Genes Genet Syst,* 90**,** 147-52.

Jinam, T. A., Kawai, Y. & Saitou, N. 2021. Modern human DNA analyses with special reference to the inner dual-structure model of Yaponesian. *Anthropological Science,* 129**,** 3-11.

Jinam, T. A., Phipps, M. E., Aghakhanian, F., Majumder, P. P., Datar, F., Stoneking, M., Sawai, H., Nishida, N., Tokunaga, K. & Kawamura, S. 2017. Discerning the origins of the Negritos, first Sundaland people: deep divergence and archaic admixture. *Genome biology and evolution,* 9**,** 2013-2022.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research,* 36**,** W5-9.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24**,** 1403-1405.

Jones, M. R. & Good, J. M. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology,* 25**,** 185-202.

Kamatani, N., Sekine, A., Kitamoto, T., Iida, A., Saito, S., Kogame, A., Inoue, E., Kawamoto, M., Harigai, M. & Nakamura, Y. 2004. Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. *American Journal of Human Genetics,* 75**,** 190-203.

Keller, A., Hempstead, M., Gomez, I. A., Gilbert, A. N. & Vosshall, L. B. 2012. An olfactory demography of a diverse metropolitan population. *BMC Neuroscience,* 13**,** 122.

Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. 2007a. Genetic variation in a human odorant receptor alters odour perception. *Nature,* 449**,** 468-72.

Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. 2007b. Genetic variation in a human odorant receptor alters odour perception. *Nature,* 449**,** 468-472.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. 2002. The human genome browser at UCSC. *Genome Research,* 12**,** 996-1006.

Knaus, B. J. & Grünwald, N. J. 2017. vcfr: a package to manipulate and visualize variant call format data in R. *Molecular ecology resources,* 17**,** 44-53.

Koganebuchi, K., Katsumura, T., Nakagome, S., Ishida, H., Kawamura, S. & Oota, H. 2012a. Autosomal and Y-chromosomal STR markers reveal a close relationship between Hokkaido Ainu and Ryukyu islanders. *Anthropological Science***,** 120322.

Koganebuchi, K., Katsumura, T., Nakagome, S., Ishida, H., Kawamura, S., Oota, H. & The Asian Archival DNA Repository Consortium 2012b. Autosomal and Y-chromosomal STR markers reveal a close relationship between Hokkaido Ainu and Ryukyu islanders. *Anthropological Science,* 120**,** 199-208.

Koganebuchi, K. & Oota, H. 2021. Paleogenomics of human remains in East Asia and Yaponesia focusing on current advances and future directions. *Anthropological Science,* 129**,** 59-69.

Korunes, K. L. & Samuk, K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular ecology resources,* 21**,** 1359-1368.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology,* 305**,** 567-80.

Lane, R. P., Cutforth, T., Young, J., Athanasiou, M., Friedman, C., Rowen, L., Evans, G., Axel, R., Hood, L. & Task, B. J. 2001. Genomic analysis of orthologous mouse and human olfactory receptor loci. *Proceedings of the National Academy of Sciences of the United States of America*, 98**,** 7390-7395.

Larsen, P. A., Heilman, A. M. & Yoder, A. D. 2014. The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genomics,* 15**,** 720.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-93.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics,* 26**,** 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

Liu, J., Deng, Y., Yu, B., Mo, B., Luo, L., Yang, J., Zhang, X., Wang, Z., Wang, Y. & Zhu, J. 2021. Targeted resequencing showing novel common and rare genetic variants increases the risk of asthma in the Chinese Han population. *Journal of clinical laboratory analysis***,** e23813.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F. & Young, N. 2013. The genotype-tissue expression (GTEx) project. *Nature genetics,* 45**,** 580-585.

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J. K. & Montgomery, S. B. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science,* 335**,** 823-828.

Mainland, J. D., Keller, A., Li, Y. R., Zhou, T., Trimmer, C., Snyder, L. L., Moberly, A. H., Adipietro, K. A., Liu, W. L. L. & Zhuang, H. 2014. The missense of smell: functional variability in the human odorant receptor repertoire. *Nature neuroscience,* 17**,** 114-120.

Mainland, J. D., Willer, J. R., Matsunami, H. & Katsanis, N. 2013. Next-generation sequencing of the human olfactory receptors. *Methods in Molecular Biology,* 1003**,** 133-47.

Majid, A. 2021. Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences,* 25**,** 111-123.

Majid, A. & Kruspe, N. 2018. Hunter-gatherer olfaction is special. *Current Biology,* 28**,** 409-413 e2.

Mashukova, A., Spehr, M., Hatt, H. & Neuhaus, E. M. 2006. β-arrestin2-mediated internalization of mammalian odorant receptors. *Journal of Neuroscience,* 26**,** 9902-9912.

Matsukusa, H., Oota, H., Haneji, K., Toma, T., Kawamura, S. & Ishida, H. 2010. A genetic analysis of the Sakishima islanders reveals no relationship with Taiwan aborigines but shared ancestry with Ainu and main-island Japanese. *American Journal of Physical Anthropology,* 142**,** 211-23.

Matsumoto, H., Miyazaki, T., Omoto, K., Misawa, S., Harada, S., Hirai, M., Sumpaico, J. S., Medado, P. M. & Ogonuki, H. 1979. Population genetic studies of the Philippine Negritos. II. gm and km allotypes of three population groups. *American Journal of Human Genetics,* 31**,** 70.

McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J. V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., Wasef, S., Shoocongdej, R., Souksavatdy, V., Sayavongkhamdy, T., Saidin, M. M., Allentoft, M. E., Sato, T., Malaspinas, A. S., Aghakhanian, F. A., Korneliussen, T., Prohaska, A., Margaryan, A., de Barros Damgaard, P., Kaewsutthi, S., Lertrit, P., Nguyen, T. M. H., Hung, H. C., Minh Tran, T., Nghia Truong, H., Nguyen, G. H., Shahidan, S., Wiradnyana, K., Matsumae, H., Shigehara, N., Yoneda, M., Ishida, H., Masuyama,

T., Yamada, Y., Tajima, A., Shibata, H., Toyoda, A., Hanihara, T., Nakagome, S., Deviese, T., Bacon, A. M., Duringer, P., Ponche, J. L., Shackelford, L., Patole-Edoumba, E., Nguyen, A. T., Bellina-Pryce, B., Galipaud, J. C., Kinaston, R., Buckley, H., Pottier, C., Rasmussen, S., Higham, T., Foley, R. A., Lahr, M. M., Orlando, L., Sikora, M., Phipps, M. E., Oota, H., Higham, C., Lambert, D. M. & Willerslev, E. 2018. The prehistoric peopling of Southeast Asia. *Science,* 361**,** 88-92.

McGann, J. P. 2017. Poor human olfaction is a 19th-century myth. *Science,* 356.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-303.

McRae, J. F., Jaeger, S. R., Bava, C. M., Beresford, M. K., Hunter, D., Jia, Y., Chheang, S. L., Jin, D., Peng, M., Gamble, J. C., Atkinson, K. R., Axten, L. G., Paisley, A. G., Williams, L., Tooman, L., Pineau, B., Rouse, S. A. & Newcomb, R. D. 2013. Identification of regions associated with variation in sensitivity to food-related odors in the human genome. *Current Biology,* 23**,** 1596-600.

McRae, J. F., Mainland, J. D., Jaeger, S. R., Adipietro, K. A., Matsunami, H. & Newcomb, R. D. 2012. Genetic variation in the odorant receptor OR2J3 is associated with the ability to detect the "grassy" smelling odor, cis-3-hexen-1-ol. *Chemical senses,* 37**,** 585-593.

Menashe, I., Abaffy, T., Hasin, Y., Goshen, S., Yahalom, V., Luetje, C. W. & Lancet, D. 2007. Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biology,* 5**,** e284.

Menashe, I., Man, O., Lancet, D. & Gilad, Y. 2002. Population differences in haplotype structure within a human olfactory receptor gene cluster. *Human Molecular Genetics,* 11**,** 1381-90.

Moore, S. F. & Puritt, P. 2017. *The Chagga and Meru of Tanzania: East Central Africa Part XVIII,* Routledge.

Nei, M. 1982. Evolution of human races at the gene level. *Human genetics, part A: The unfolding genome***,** 167-181.

Nei, M. & Kumar, S. 2000. *Molecular evolution and phylogenetics,* Oxford university press.

Nei, M. & Li, W.-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences,* 76**,** 5269-5273.

Niimura, Y. 2013. Identification of olfactory receptor genes from mammalian genome sequences. *Methods in Molecular Biology,* 1003**,** 39-49.

Niimura, Y., Matsui, A. & Touhara, K. 2018. Acceleration of olfactory receptor gene loss in primate evolution: possible link to anatomical change in sensory systems and dietary transition. *Molecular Biology and Evolution,* 35**,** 1437–1450.

Niimura, Y. & Nei, M. 2003. Evolution of olfactory receptor genes in the human genome. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 12235-12240.

Nishida, N., Hirai, M., Kawamura, S., Oota, H., Umetsu, K., Kimura, R., Ohashi, J., Tajima, A., Yamamoto, T. & Tanabe10, H. 2012. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *Journal of human genetics,* 1**,** 9.

Nishimoto, K., Kochi, Y., Ikari, K., Yamamoto, K., Suzuki, A., Shimane, K., Nakamura, Y., Yano, K., Iikuni, N., Tsukahara, S., Kamatani, N., Okamoto, H., Kaneko, H., Kawaguchi, Y., Hara, M., Toyama, Y., Horiuchi, T., Tao, K., Yasutomo, K., Hamada, D., Yasui, N., Inoue, H., Itakura, M., Yamanaka, H. & Momohara, S. 2010. Association study of *TRAF1-C5* polymorphisms with susceptibility to rheumatoid

arthritis and systemic lupus erythematosus in Japanese. *Annals of the Rheumatic Diseases,* 69**,** 368-373.

North, K. E., Martin, L. J. & Crawford, M. H. 2000. The origins of the Irish travellers and the genetic structure of Ireland. *Annals of Human Biology,* 27**,** 453-465.

Nozawa, M., Kawahara, Y. & Nei, M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. *Proceedings of the National Academy of Sciences,* 104**,** 20421-20426.

Olender, T., Nativ, N. & Lancet, D. 2013. HORDE: comprehensive resource for olfactory receptor genomics. *Olfactory Receptors.* Springer.

Olofsson, J. K. & Wilson, D. A. 2018. Human olfaction: it takes two villages. *Current Biology,* 28**,** R108-R110.

Omoto, K. 1979. Carbonic anhydrase-I polymorphism in a Philippine aboriginal population. *American journal of human genetics,* 31**,** 747.

Omoto, K. 1980. Genetic variants of red cell enzymes as potential anthropological markers in the Western Pacific. *Hemoglobin,* 4**,** 755-760.

Omoto, K. 1981. The genetic origins of the Philippine Negritos. *Current Anthropology,* 22**,** 421-422.

Omoto, K. 1982. Allelic diversity of human populations in the Asian-Pacific area. *Molecular Evolution, Protein Polymorphism and the Neural Theory***,** 193-214.

Omoto, K. 1984. The negritos: Genetic origins and microevolution. *Acta anthropogenetica,* 8**,** 137-147.

Omoto, K., Misawa, S., Harada, S., Sumpaico, J. S., Medado, P. M. & Ogonuki, H. 1978. Population genetic studies of the Philippine Negritos. I. A pilot survey of red cell enzyme and serum protein groups. *American journal of human genetics,* 30**,** 190.

Omoto, K., Ueda, S., Goriki, K., Takahashi, N., Misawa, S. & Pagaran, I. G. 1981. Population genetic studies of the Philippine Negritos. III. Identification of the carbonic anhydrase-1 variant with CA1 Guam. *American Journal of Human Genetics,* 33**,** 105.

Osada, N. & Kawai, Y. 2021. Exploring models of human migration to the Japanese archipelago using genome-wide genetic data. *Anthropological Science,* 129**,** 45-58.

Osier, M. V., Cheung, K.-H., Kidd, J. R., Pakstis, A. J., Miller, P. L. & Kidd, K. K. 2001. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms—an update. *Nucleic acids research,* 29**,** 317-319.

Osier, M. V., Cheung, K. H., Kidd, J. R., Pakstis, A. J., Miller, P. L. & Kidd, K. K. 2002. ALFRED: an allele frequency database for anthropology. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists,* 119**,** 77-83.

Ozeki, T., Furuya, Y., Nagano, C., Matsui, C., Takayanagi, R., Yokoyama, H. & Yamada, Y. 2006. Analysis of linkage between lymphotoxin alpha haplotype and polymorphisms in 5'-flanking region of tumor necrosis factor alpha gene associated with efficacy of infliximab for Crohn's disease patients. *Mutation Research,* 602**,** 170-174.

Pembleton, L. W., Cogan, N. O. & Forster, J. W. 2013. St AMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular ecology resources,* 13**,** 946-952.

Pluznick, J. L., Zou, D.-J., Zhang, X., Yan, Q., Rodriguez-Gil, D. J., Eisner, C., Wells, E., Greer, C. A., Wang, T. & Firestein, S. 2009. Functional expression of the olfactory signaling system in the kidney. *Proceedings of the National Academy of Sciences,* 106**,** 2059-2064.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A. & Roazen, D. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.

Prabhu, S. & Pe'Er, I. 2009. Overlapping pools for high-throughput targeted resequencing. *Genome research,* 19**,** 1254-1261.

Putnam, P. 1948. The pygmies of the Ituri forest. *A reader in general anthropology***,** 322-342.

Quinlan, A. R. & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics,* 16**,** 276-7.

Roberts, S. C., Havlicek, J. & Schaal, B. 2020. Human olfactory communication: current challenges and future prospects. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 375.

Rouquier, S., Blancher, A. & Giorgi, D. 2000. The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. *Proceedings of the National Academy of Sciences,* 97**,** 2870-2874.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. 2009. Odor coding by a Mammalian receptor repertoire. *Science signaling,* 2**,** ra9-ra9.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution,* 4**,** 406-425.

Sakamoto, Y., Inoue, H., Keshavarz, P., Miyawaki, K., Yamaguchi, Y., Moritani, M., Kunika, K., Nakamura, N., Yoshikawa, T., Yasui, N., Shiota, H., Tanahashi, T. & Itakura, M. 2007. SNPs in the *KCNJ11-ABCC8* gene locus are associated with type 2

diabetes and blood pressure levels in the Japanese population. *Journal of Human Genetics,* 52**,** 781-793.

Sanz, G., Schlegel, C., Pernollet, J.-C. & Briand, L. 2005. Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chemical senses,* 30**,** 69-80.

Sato, T., Nakagome, S., Watanabe, C., Yamaguchi, K., Kawaguchi, A., Koganebuchi, K., Haneji, K., Yamaguchi, T., Hanihara, T., Yamamoto, K., Ishida, H., Mano, S., Kimura, R. & Oota, H. 2014. Genome-wide SNP analysis reveals population structure and demographic history of the ryukyu islanders in the southern part of the Japanese archipelago. *Molecular Biology and Evolution,* 31**,** 2929-40.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S. & Klimke, W. 2021. Database resources of the national center for biotechnology information. *Nucleic acids research,* 49**,** D10.

Schmiedeberg, K., Shirokova, E., Weber, H.-P., Schilling, B., Meyerhof, W. & Krautwurst, D. 2007. Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *Journal of structural biology,* 159**,** 400-412.

Schmieder, R. & Edwards, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics,* 27**,** 863-4.

Schweizer, R. M., Robinson, J., Harrigan, R., Silva, P., Galverni, M., Musiani, M., Green, R. E., Novembre, J. & Wayne, R. K. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology,* 25**,** 357-79.

Secundo, L., Snitz, K., Weissler, K., Pinchover, L., Shoenfeld, Y., Loewenthal, R., Agmon-Levin, N., Frumin, I., Bar-Zvi, D., Shushan, S. & Sobel, N. 2015. Individual olfactory

perception reveals meaningful nonolfactory genetic information. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 8750-5.

Serres, M. 2008. *The five senses: a philosophy of mingled bodies,* Bloomsbury Publishing.

Shepard, B. D. 2021. The Sniffing Kidney: Roles for Renal Olfactory Receptors in Health and Disease. *Kidney360*.

Shirasu, M., Yoshikawa, K., Takai, Y., Nakashima, A., Takeuchi, H., Sakano, H. & Touhara, K. 2014. Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron,* 81**,** 165-178.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research,* 19**,** 1117-23.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics,* 15**,** 121-32.

Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J. A., Wetzel, C. H., Zimmer, R. K. & Hatt, H. 2003. Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science,* 299**,** 2054-2058.

Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S., Omoto, K. & Horai, S. 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of human genetics,* 49**,** 187-193.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* 123**,** 585-95.

Takata, Y., Inoue, H., Sato, A., Tsugawa, K., Miyatake, K., Hamada, D., Shinomiya, F., Nakano, S., Yasui, N., Tanahashi, T. & Itakura, M. 2008. Replication of reported genetic associations of *PADI4*, *FCRL3*, *SLC22A4* and *RUNX1* genes with rheumatoid

arthritis: results of an independent Japanese population and evidence from meta-analysis of East Asian studies. *Journal of Human Genetics,* 53**,** 163-173.

Thurston, T. L. 1997. Historians, prehistorians, and the tyranny of the historical record: Danish state formation through documents and archaeological data. *Journal of Archaeological Method and Theory,* 4**,** 239-263.

Trimmer, C., Keller, A., Murphy, N. R., Snyder, L. L., Willer, J. R., Nagai, M. H., Katsanis, N., Vosshall, L. B., Matsunami, H. & Mainland, J. D. 2019. Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proceedings of the National Academy of Sciences of the United States of America,* 116**,** 9475-9480.

Turnbull, C. 1965. The Mbuti Pygmies: an ethnographic survey. *Anthropological Papers of the American Museum of Natural History,* 50**,** 139-282.

Turnbull, C. M. 1983. *The Mbuti Pygmies: change and adaptation*, Case Studies in Cultural Anthr.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S. & DePristo, M. A. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics,* 43**,** 11 10 1-11 10 33.

Verbeurgt, C., Wilkin, F., Tarabichi, M., Gregoire, F., Dumont, J. E. & Chatelain, P. 2014. Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PloS one,* 9**,** e96333.

Wallis, E. E. 1987. The dynamics of vocalic harmony in Shapsugh Circassian. *Word,* 38**,** 81-90.

Watanabe, Y., Isshiki, M. & Ohashi, J. 2021. Prefecture-level population structure of the Japanese based on SNP genotypes of 11,069 individuals. *Journal of Human Genetics,* 66**,** 431-437.

Wetzel, C. H., Oles, M., Wellerdieck, C., Kuczkowiak, M., Gisselmann, G. & Hatt, H. 1999. Specificity and sensitivity of a human olfactory receptor functionally expressed in human embryonic kidney 293 cells andxenopus laevis oocytes. *Journal of Neuroscience,* 19**,** 7426-7433.

Wooding, S. 2013. Olfaction: it makes a world of scents. *Current Biology,* 23**,** R677-9.

Yohe, L. R., Davies, K. T. J., Simmons, N. B., Sears, K. E., Dumont, E. R., Rossiter, S. J. & Davalos, L. M. 2020. Evaluating the performance of targeted sequence capture, RNA-Seq, and degenerate-primer PCR cloning for sequencing the largest mammalian multigene family. *Molecular Ecology Resources,* 20**,** 140-153.

Zhang, W., Zhu, G., Tan, L., Yang, J., Chen, Y., Liu, Q., Shen, Q., Chen, J. & Zhang, L. 2014. Role of olfaction in the foraging behavior and trial-and-error learning in short-nosed fruit bat, Cynopterus sphinx. *Behavioural processes,* 103**,** 23-27.

Zhang, X. & Firestein, S. 2002. The olfactory receptor gene superfamily of the mouse. *Nature Neuroscience,* 5**,** 124-33.

Zheng-Bradley, X., Streeter, I., Fairley, S., Richardson, D., Clarke, L., Flicek, P. & Genomes Project, C. 2017. Alignment of 1000 Genomes Project reads to reference assembly GRCh38. *GigaScience,* 6**,** 1-8.

Zhuang, H. Y., Chien, M. S. & Matsunami, H. 2009. Dynamic functional evolution of an odorant receptor for sex-steroid-derived odors in primates. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 21247-21251.

APPENDICES

Appendix I: List of ORs with known ligands			
Gene Name from Niimura et al. 2018 Mol Biol Evol 35:1437	**HORDE name**	**Agonist**	**Reference**
HsOR1.4.10	OR10Z1	Diacetyl	(Trimmer et al., 2019)
HsOR1.4.10	OR10Z1	2−Butanone	(Trimmer et al., 2019)
HsOR1.4.27	OR10J5	Lyral	(Saito et al., 2009)
HsOR1.4.27	OR10J5	Bourgeonal	(Trimmer et al., 2019)
HsOR1.4.5	OR10R2	Diacetyl	(Trimmer et al., 2019)
HsOR1.4.5	OR10R2	2−Butanone	(Trimmer et al., 2019)
HsOR1.4.7	OR6Y1	Diacetyl	(Trimmer et al., 2019)
HsOR1.4.7	OR6Y1	2−Butanone	(Trimmer et al., 2019)
HsOR1.4.8	OR6P1	Anisaldehyde	(Mainland et al., 2014)
HsOR1.5.1	OR2B11	Coumarin	(Adipietro et al., 2012)
HsOR1.5.10	OR1C1	Linalool	(Adipietro et al., 2012)
HsOR1.5.10	OR1C1	Linalool	(Trimmer et al., 2019)
HsOR1.5.15	OR2W3	4−Methylvaleric acid	(Trimmer et al., 2019)
HsOR1.5.35	OR2M7	Citronellol	(Saito et al., 2009)
HsOR1.5.46	OR2T10	Nonyl aldehyde	(Trimmer et al., 2019)
HsOR1.5.6b	OR13G1	Butyric acid	(Trimmer et al., 2019)
HsOR2.4.1	OR6B2	Isobutyraldehyde	(Trimmer et al., 2019)
HsOR22.1.1	OR11H1	Isovaleric acid	(Trimmer et al., 2019)

HsOR3.3.16	OR5K1	Eugenol methyl ether	(Adipietro et al., 2012)
HsOR5.4.2	OR2Y1	Heptyl acetate	(Trimmer et al., 2019)
HsOR6.3.16	OR14J1	Caproic acid	(Trimmer et al., 2019)
HsOR6.3.19	OR12D2	2−Ethylfenchol	(Trimmer et al., 2019)
HsOR6.3.2	OR2W1	1-octanol	(Saito et al., 2009)
HsOR6.3.21	OR11A1	2-ethyl fenchol	(Adipietro et al., 2012)
HsOR6.3.21	OR11A1	2−Ethylfenchol	(Trimmer et al., 2019)
HsOR6.3.21	OR11A1	Fenchone	(Trimmer et al., 2019)
HsOR6.3.22	OR10C1	2−Ethylfenchol	(Trimmer et al., 2019)
HsOR6.3.6	OR2J3	Cis-3-hexen-1-ol	(McRae et al., 2012)
HsOR6.3.8	OR2J2	1-octanol	(Saito et al., 2009)
HsOR7.6.11	OR2A25	Geranyl acetate	(Adipietro et al., 2012)
HsOR7.6.11	OR2A25	Linalool	(Trimmer et al., 2019)
HsOR9.4.4	OR13C8	Citral	(Trimmer et al., 2019)
HsOR9.6.14	OR5C1	Citral	(Trimmer et al., 2019)
HsOR11.11.15	OR4C15	Paraffin oil	(Trimmer et al., 2019)
HsOR11.11.29	OR5D18	Eugenol	(Braun et al., 2007)
HsOR11.11.3	OR4A16	Eugenol methyl ether	(Trimmer et al., 2019)
HsOR11.11.34	OR5W2	(−)−Menthol	(Trimmer et al., 2019)
HsOR11.11.59	OR8K3	(+)-menthol	(Adipietro et al., 2012)
HsOR11.12.11	OR10Q1	2−Ethylfenchol	(Trimmer et al., 2019)
HsOR11.13.11	OR4D9	b-ionone	(Jaeger et al., 2013a)
HsOR11.13.3	OR5AN1	Muscone	(Shirasu et al., 2014)
HsOR11.13.5	OR5A2	b-ionone	(Jaeger et al., 2013a)

HsOR11.13.5	OR5A2	Galaxolide	(Trimmer et al., 2019)
HsOR11.13.6	OR5A1	b-ionone	(Jaeger et al., 2013a)
HsOR11.13.7	OR4D6	b-ionone	(Jaeger et al., 2013a)
HsOR11.13.7	OR4D6	Galaxolide	(Trimmer et al., 2019)
HsOR11.18.11	OR10G4	Ethyl vanillin	(Adipietro et al., 2012)
HsOR11.18.11	OR10G4	Guaiacol	(Trimmer et al., 2019)
HsOR11.18.12	OR10G9	Ethyl vanillin	(Adipietro et al., 2012)
HsOR11.18.12	OR10G9	Isoeugenol	(Trimmer et al., 2019)
HsOR11.18.14	OR10G7	Eugenol	(Adipietro et al., 2012)
HsOR11.18.19	OR10D3	Isoeugenol	(Trimmer et al., 2019)
HsOR11.18.26	OR8D1	4,5-dimethyl-3-hydroxy-2,5-dihydrofuran-2-one	(Adipietro et al., 2012)
HsOR11.18.34	OR8B3	(+)-carvone	(Mainland et al., 2014)
HsOR11.18.5	OR8D4	Guaiacol	(Trimmer et al., 2019)
HsOR11.18.5	OR8D4	Isoeugenol	(Trimmer et al., 2019)
HsOR11.18.6	OR4D5	r−Limonene	(Trimmer et al., 2019)
HsOR11.3.14	OR51E1	Nonanoic acid	(Fujita et al., 2007)
HsOR11.3.16	OR51E2	Propionic acid	(Saito et al., 2009)
HsOR11.3.40	OR51L1	4-allylphenylacetate	(Saito et al., 2009)
HsOR11.3.49P*	OR52A4	Octyl aldehyde	(Trimmer et al., 2019)
HsOR11.3.68	OR52D1	Ethyl isobutyrate	(Sanz et al., 2005)
HsOR11.3.93	OR56A4	Isovaleric acid	(Trimmer et al., 2019)
HsOR11.3.93	OR56A4	Isobutyric acid	(Trimmer et al., 2019)
HsOR11.3.94	OR56A1	Isovaleric acid	(Trimmer et al., 2019)

HsOR11.4.2	OR2AG1	Amyl butyrate	(Mashukova et al., 2006)
HsOR11.4.3	OR6A2	Cineole	(Trimmer et al., 2019)
HsOR11.5.5	OR5P3	(+)-carvone	(Saito et al., 2009)
HsOR11.5.7	OR10A6	3-phenyl propyl propionate	(Mainland et al., 2014)
HsOR12.5.11	OR6C1	Heptyl acetate	(Trimmer et al., 2019)
HsOR14.1.15	OR4K15	Diallyl sulfide	(Trimmer et al., 2019)
HsOR14.1.27	OR11H6	Isovaleric acid	(Menashe et al., 2007)
HsOR14.1.28P0	OR11H7P	Isovaleric acid	(Menashe et al., 2007)
HsOR14.1.29	OR11H4	Isovaleric acid	(Menashe et al., 2007)
HsOR14.1.3	OR4Q3	Eugenol	(Mainland et al., 2014)
HsOR14.2.2	OR10G3	Ethyl vanillin	(Adipietro et al., 2012)
HsOR14.2.5	OR4E2	Amyl acetate	(Mainland et al., 2014)
HsOR15.2.2	OR4F15	Butyl acetate	(Trimmer et al., 2019)
HsOR16.1.3	OR2C1	Octanethiol	(Saito et al., 2009)
HsOR17.1.10	OR3A2	Androstadienone	(Trimmer et al., 2019)
HsOR17.1.11	OR3A1	Helional	(Wetzel et al., 1999)
HsOR17.1.2	OR1D2	Bourgeonal	(Spehr et al., 2003)
HsOR17.1.4	OR1G1	1-nonanol	(Sanz et al., 2005)
HsOR17.1.6	OR1A2	Citronellal	(Schmiedeberg et al., 2007)
HsOR17.1.7	OR1A1	Dihydrojasmone	(Saito et al., 2009)
HsOR17.1.7	OR1A1	Caproic acid	(Trimmer et al., 2019)
HsOR19.2.11	OR7D4	Androstenone	(Keller et al., 2007b)

HsOR19.2.11	OR7D4	Androstenone	(Trimmer et al., 2019)
HsOR19.2.11	OR7D4	Androstadienone	(Trimmer et al., 2019)
HsOR19.3.1	OR7C1	Androstadienone	(Mainland et al., 2014)
*Not included in this study			

Appendix II. A list of OR genes considered in this study.							
Gene No. in This Study	Gene Name from Niimura et al. 2018 Mol Biol Evol 35:1437	Gene Name from HORDE [a]	Chromsome / Scaffold	Gene Position Coordinate in Database [b]	Direction	Category [c]	Probe Design Length
1	HsOR1.1.2	OR4G11P	chr1	63015-63887	+	F	1073
2	HsOR1.1.3	OR4F5	chr1	69091-70008	+	F	1118
3	HsOR1.1.4	OR4F29	chr1	450740-451678	-	F	1139
4	HsOR1.1.5	OR4F16	chr1	685716-686654	-	F	1139
5	HsOR1.4.1	OR10T2	chr1	158398522-158399466	-	F	1145
6	HsOR1.4.2	OR10K2	chr1	158419928-158420866	-	F	1139
7	HsOR1.4.4	OR10K1	chr1	158465562-158466503	+	F	1142
8	HsOR1.4.5	OR10R2	chr1	158479911-158480885	+	F	1175
9	HsOR1.4.7	OR6Y1	chr1	158547128-158548105	-	F	1178
10	HsOR1.4.8	OR6P1	chr1	158562651-158563604	-	F	1154
11	HsOR1.4.9	OR10X1	chr1	158578919-158579848	-	F	1130
12	HsOR1.4.10	OR10Z1	chr1	158606439-158607380	+	F	1142
13	HsOR1.4.12	OR6K2	chr1	158699678-158700652	-	F	1175
14	HsOR1.4.13	OR6K3	chr1	158717168-158718115	-	F	1148
15	HsOR1.4.16	OR6K6	chr1	158754888-158755847	+	F	1160
16	HsOR1.4.17	OR6N1	chr1	158765744-158766682	-	F	1139

17	HsOR1.4.18	OR6N2	chr1	158776682-158777635	-	F	1154
18	HsOR1.4.21	OR10J3	chr1	159313670-159314659	-	F	1190
19	HsOR1.4.26	OR10J1	chr1	159439759-159440721	+	F	1163
20	HsOR1.4.27	OR10J5	chr1	159535078-159536007	-	F	1130
21	HsOR1.5.1	OR2B11	chr1	247451029-247451982	-	F	1154
22	HsOR1.5.3	OR2C3	chr1	247531549-247532511	-	F	1163
23	HsOR1.5.4	OR2G2	chr1	247588360-247589313	+	F	1154
24	HsOR1.5.5	OR2G3	chr1	247605586-247606515	+	F	1130
25	HsOR1.5.6	OR14L1P	chr1	247619681-247620562	+	F	1082
26	HsOR1.5.6b	OR13G1	chr1	247672118-247673041	-	F	1124
27	HsOR1.5.7	OR6F1	chr1	247711829-247712755	-	F	1127
28	HsOR1.5.8	OR14A2	chr1	247723099-247724043	-	F	1145
29	HsOR1.5.9	OR14K1	chr1	247738615-247739559	+	F	1145
30	HsOR1.5.10	OR1C1	chr1	247757462-247758406	-	F	1145
31	HsOR1.5.12	OR14A16	chr1	247814800-247815729	-	F	1130
32	HsOR1.5.14	OR11L1	chr1	247840928-247841896	-	F	1169
33	HsOR1.5.15	OR2W3	chr1	247895587-247896531	+	F	1145

34	HsOR1.5.16	OR2T8	chr1	247921018-247921956	+	F	1139
35	HsOR1.5.17	OR2AJ1	chr1	247933766-247934755	+	F	1190
36	HsOR1.5.18	OR2L8	chr1	247948858-247949796	+	F	1139
37	HsOR1.5.19	OR2AK2	chr1	247965377-247966339	+	F	1163
38	HsOR1.5.23	OR2L5	chr1	248021948-248022886	+	F	1139
39	HsOR1.5.24	OR2L2	chr1	248038268-248039206	+	F	1139
40	HsOR1.5.25	OR2L3	chr1	248060682-248061620	+	F	1139
41	HsOR1.5.27	OR2L13	chr1	248099376-248100314	+	F	1139
42	HsOR1.5.29	OR2M5	chr1	248145148-248146086	+	F	1139
43	HsOR1.5.30	OR2M2	chr1	248179986-248181029	+	F	1244
44	HsOR1.5.31	OR2M3	chr1	248203068-248204006	+	F	1139
45	HsOR1.5.32	OR2M4	chr1	248238929-248239864	+	F	1136
46	HsOR1.5.33	OR2T33	chr1	248272852-248273814	-	F	1163
47	HsOR1.5.34	OR2T12	chr1	248294616-248295578	-	F	1163
48	HsOR1.5.35	OR2M7	chr1	248323630-248324568	-	F	1139
49	HsOR1.5.36	OR14C36	chr1	248348775-248349713	+	F	1139
50	HsOR1.5.37	OR2T4	chr1	248361665-248362627	+	F	1163

51	HsOR1.5.38	OR2T6	chr1	248387609-248388535	+	F	1127
52	HsOR1.5.39	OR2T1	chr1	248406148-248407104	+	F	1157
53	HsOR1.5.40	OR2T7	chr1	248441117-248442133	+	F	1217
54	HsOR1.5.41	OR2T2	chr1	248452798-248453772	+	F	1175
55	HsOR1.5.42	OR2T3	chr1	248473351-248474307	+	F	1157
56	HsOR1.5.43	OR2T5	chr1	248488589-248489536	+	F	1148
57	HsOR1.5.43a	OR2G6	chr1	248521647-248522597	+	F	1151
58	HsOR1.5.44	OR2T29	chr1	248558544-248559491	-	F	1148
59	HsOR1.5.45	OR2T34	chr1	248573801-248574757	-	F	1157
60	HsOR1.5.46	OR2T10	chr1	248592830-248593768	-	F	1139
61	HsOR1.5.47	OR2T11	chr1	248626178-248627128	-	F	1151
62	HsOR1.5.48	OR2T35	chr1	248638287-248639258	-	F	1172
63	HsOR1.5.49	OR2T27	chr1	248649931-248650884	-	F	1154
64	HsOR1.5.50	OR14I1	chr1	248681369-248682304	-	F	1136
65	HsOR2.4.1	OR6B2	chr2	240029491-240030429	-	F	1139
66	HsOR2.4.2	OR6B3	chr2	240045077-240046072	-	F	1196
67	HsOR3.3.2	OR5AC2	chr3	98087173-98088102	+	F	1130

68	HsOR3.3.4	OR5H1	chr3	98132698-98133639	+	F	1142
69	HsOR3.3.5	OR5H14	chr3	98149386-98150318	+	F	1133
70	HsOR3.3.6	OR5H15	chr3	98168700-98169641	+	F	1142
71	HsOR3.3.11	OR5H6	chr3	98264333-98265262	+	F	1130
72	HsOR3.3.12	OR5H2	chr3	98282888-98283832	+	F	1145
73	HsOR3.3.14	OR5K4	chr3	98353854-98354819	+	F	1166
74	HsOR3.3.15	OR5K3	chr3	98390666-98391631	+	F	1166
75	HsOR3.3.16	OR5K1	chr3	98469577-98470503	+	F	1127
76	HsOR3.3.17	OR5K2	chr3	98497681-98498631	+	F	1151
77	HsOR5.4.2	OR2Y1	chr5	180739123-180740058	-	F	1136
78	HsOR5.4.3	OR2V1	chr5	181124357-181125304	-	F	1148
79	HsOR5.4.4	OR2V2	chr5	181154943-181155890	+	F	1148
80	HsOR5.4.5	OR4F3	chr5	181367287-181368225	+	F	1139
81	HsOR6.2.1	OR2B2	chr6	27911246-27912319	-	F	1274
82	HsOR6.2.3	OR2B6	chr6	27957241-27958182	+	F	1142
83	HsOR6.2.8	OR1F12	chr6	28073316-28074329	+	F	1214
84	HsOR6.3.2	OR2W1	chr6	29044213-29045175	-	F	1163

85	HsOR6.3.4	OR2B3	chr6	29086307-29087248	-	F	1142
86	HsOR6.3.5	OR2J1P	chr6	29100943-29101881	+	F	1139
87	HsOR6.3.6	OR2J3	chr6	29111891-29112826	+	F	1136
88	HsOR6.3.8	OR2J2	chr6	29173636-29174574	+	F	1139
89	HsOR6.3.16	OR14J1	chr6	29306690-29307655	+	F	1166
90	HsOR6.3.17	OR5V1	chr6	29355230-29356195	-	F	1166
91	HsOR6.3.18	OR12D3	chr6	29374337-29375287	-	F	1151
92	HsOR6.3.19	OR12D2	chr6	29396700-29397623	+	F	1124
93	HsOR6.3.21	OR11A1	chr6	29426694-29427641	-	F	1148
94	HsOR6.3.22	OR10C1	chr6	29440016-29440954	+	F	1139
95	HsOR6.3.23	OR2H1	chr6	29461770-29462720	+	F	1151
96	HsOR6.3.26	OR2H2	chr6	29587945-29588883	+	F	1139
97	HsOR6.4.1	OR2A4	chr6	131700469-131701401	-	F	1133
98	HsOR7.4.1	OR2AE1	chr7	99876062-99877033	-	F	1172
99	HsOR7.5.3	OR9A4	chr7	141918876-141919820	+	F	1145
100	HsOR7.6.2	OR6V1	chr7	143052317-143053282	+	F	1166
101	HsOR7.6.6	OR2F2	chr7	143935233-143936186	+	F	1154

102	HsOR7.6.7	OR2F1	chr7	143959971-143960924	+	F	1154
103	HsOR7.6.9	OR6B1	chr7	144003997-144004932	+	F	1136
104	HsOR7.6.10	OR2A5	chr7	144050402-144051337	+	F	1136
105	HsOR7.6.11	OR2A25	chr7	144074220-144075152	+	F	1133
106	HsOR7.6.12	OR2A12	chr7	144095108-144096040	+	F	1133
107	HsOR7.6.13	OR2A2	chr7	144109583-144110539	+	F	1157
108	HsOR7.6.15	OR2A14	chr7	144129113-144130045	+	F	1133
109	HsOR7.6.19	OR2A42	chr7	144231911-144232843	-	F	1133
110	HsOR7.6.21	OR2A7	chr7	144258696-144259628	-	F	1133
111	HsOR7.6.23	OR2A1	chr7	144318125-144319057	+	F	1133
112	HsOR8.1.1	OR4F21	chr8	166086-167024	-	F	1139
113	HsOR9.1.2	OR13J1	chr9	35869463-35870401	-	F	1139
114	HsOR9.1.3	OR2S2	chr9	35957139-35958098	-	F	1160
115	HsOR9.4.1	OR13F1	chr9	104504263-104505222	+	F	1160
116	HsOR9.4.2	OR13C4	chr9	104526253-104527209	-	F	1157
117	HsOR9.4.3	OR13C3	chr9	104535770-104536723	-	F	1154
118	HsOR9.4.4	OR13C8	chr9	104569168-104570130	+	F	1163
119	HsOR9.4.6	OR13C5	chr9	104598457-104599413	-	F	1157

120	HsOR9.4.7	OR13C2	chr9	104604671-104605627	-	F	1157
121	HsOR9.4.8	OR13C9	chr9	104617248-104618204	-	F	1157
122	HsOR9.4.11	OR13D1	chr9	104694518-104695462	+	F	1145
123	HsOR9.5.1	OR2K2	chr9	111327483-111328433	-	F	1151
124	HsOR9.6.1	OR1J1	chr9	122476958-122477926	-	F	1169
125	HsOR9.6.2	OR1J2	chr9	122510802-122511743	+	F	1142
126	HsOR9.6.3	OR1J4	chr9	122519141-122520082	+	F	1142
127	HsOR9.6.4	OR1N1	chr9	122526358-122527293	-	F	1136
128	HsOR9.6.5	OR1N2	chr9	122553212-122554162	+	F	1151
129	HsOR9.6.6	OR1L8	chr9	122567548-122568477	-	F	1130
130	HsOR9.6.8	OR1Q1	chr9	122614738-122615682	+	F	1145
131	HsOR9.6.9	OR1B1	chr9	122628579-122629535	-	F	1157
132	HsOR9.6.10	OR1L1	chr9	122661716-122662648	+	F	1133
133	HsOR9.6.11	OR1L3	chr9	122675130-122676104	+	F	1175
134	HsOR9.6.12	OR1L4	chr9	122723990-122724925	+	F	1136
135	HsOR9.6.13	OR1L6	chr9	122749848-122750783	+	F	1136
136	HsOR9.6.14	OR5C1	chr9	122788933-122789895	+	F	1163

137	HsOR9.6.15	OR1K1	chr9	122800123-122801073	+	F	1151
138	HsOR10.2.2	OR13A1	chr10	45303436-45304365	-	F	1130
139	HsOR11.3.2	OR52B4	chr11	4367351-4368295	-	F	1145
140	HsOR11.3.5	OR52P2P	chr11	4431330-4432229	-	F	1100
141	HsOR11.3.6	OR52K2	chr11	4449340-4450284	+	F	1145
142	HsOR11.3.8	OR52K1	chr11	4488901-4489845	+	F	1145
143	HsOR11.3.10	OR52M1	chr11	4545191-4546144	+	F	1154
144	HsOR11.3.11	OR52I2	chr11	4586891-4587865	+	F	1175
145	HsOR11.3.12	OR52I1	chr11	4594039-4595013	+	F	1175
146	HsOR11.3.13	OR51D1	chr11	4639791-4640765	+	F	1175
147	HsOR11.3.14	OR51E1	chr11	4652527-4653483	+	F	1157
148	HsOR11.3.16	OR51E2	chr11	4681749-4682711	-	F	1163
149	HsOR11.3.18	OR51F5P	chr11	4709569-4710468	+	F	1100
150	HsOR11.3.22	OR51F1	chr11	4768979-4769938	-	F	1160
151	HsOR11.3.24	OR52R1	chr11	4803433-4804380	-	F	1148
152	HsOR11.3.25	OR51F2	chr11	4821422-4822414	+	F	1193
153	HsOR11.3.27	OR51S1	chr11	4848237-4849208	-	F	1172

154	HsOR11.3.28	OR51H1P	chr11	4859656-4860564	-	F	1109
155	HsOR11.3.30	OR51T1	chr11	4881900-4882883	+	F	1184
156	HsOR11.3.33	OR51A7	chr11	4907370-4908308	+	F	1139
157	HsOR11.3.34	OR51G2	chr11	4914719-4915663	-	F	1145
158	HsOR11.3.35	OR51G1	chr11	4923374-4924339	-	F	1166
159	HsOR11.3.37	OR51A4	chr11	4946159-4947100	-	F	1142
160	HsOR11.3.38	OR51A2	chr11	4954772-4955713	-	F	1142
161	HsOR11.3.40	OR51L1	chr11	4998983-4999930	+	F	1148
162	HsOR11.3.43	OR52J3	chr11	5046526-5047461	+	F	1136
163	HsOR11.3.44	OR52E2	chr11	5058650-5059627	-	F	1178
164	HsOR11.3.50	OR52A5	chr11	5131692-5132642	-	F	1151
165	HsOR11.3.51	OR52A1	chr11	5151431-5152369	-	F	1139
166	HsOR11.3.52	OR51A1P	chr11	5170351-5171223	-	F	1073
167	HsOR11.3.54	OR51V1	chr11	5199735-5200700	-	F	1166
168	HsOR11.3.55	OR51B4	chr11	5301014-5301946	-	F	1133
169	HsOR11.3.57	OR51B2	chr11	5323359-5324297	-	F	1139
170	HsOR11.3.59	OR51B5	chr11	5342586-5343524	-	F	1139

171	HsOR11.3.60	OR51B6	chr11	5351508-5352446	+	F	1139
172	HsOR11.3.61	OR51M1	chr11	5389432-5390379	+	F	1148
173	HsOR11.3.63	OR51Q1	chr11	5422201-5423154	+	F	1154
174	HsOR11.3.65	OR51I1	chr11	5440570-5441514	-	F	1145
175	HsOR11.3.66	OR51I2	chr11	5453489-5454427	+	F	1139
176	HsOR11.3.68	OR52D1	chr11	5488707-5489663	+	F	1157
177	HsOR11.3.70	OR52H1	chr11	5544561-5545523	-	F	1163
178	HsOR11.3.74	OR52B6	chr11	5580907-5581884	+	F	1178
179	HsOR11.3.77	OR56B1	chr11	5736517-5737491	+	F	1175
180	HsOR11.3.78	OR52N4	chr11	5754741-5755706	+	F	1166
181	HsOR11.3.79	OR56B2P	chr11	5765162-5766049	+	F	1088
182	HsOR11.3.80	OR52N5	chr11	5777660-5778634	-	F	1175
183	HsOR11.3.81	OR52N1	chr11	5787854-5788816	-	F	1163
184	HsOR11.3.83	OR52N2	chr11	5820336-5821301	+	F	1166
185	HsOR11.3.84	OR52E6	chr11	5840956-5841897	-	F	1142
186	HsOR11.3.85	OR52E8	chr11	5856749-5857702	-	F	1154
187	HsOR11.3.87	OR52E4	chr11	5884293-5885231	+	F	1139

188	HsOR11.3.88	OR52E5	chr11	5900765-5901760	+	F	1196
189	HsOR11.3.90	OR56A3	chr11	5947347-5948294	+	F	1148
190	HsOR11.3.91	OR56A5	chr11	5967553-5968494	-	F	1142
191	HsOR11.3.92	OR52L1	chr11	5985941-5986885	-	F	1145
192	HsOR11.3.93	OR56A4	chr11	6002051-6002992	-	F	1142
193	HsOR11.3.94	OR56A1	chr11	6026748-6027704	-	F	1157
194	HsOR11.3.96	OR52L2P	chr11	6057316-6058230	-	F	1115
195	HsOR11.3.98	OR56B4	chr11	6107779-6108738	+	F	1160
196	HsOR11.3.101	OR52B2	chr11	6169355-6170326	-	F	1172
197	HsOR11.3.102	OR52W1	chr11	6199224-6200186	+	F	1163
198	HsOR11.4.1	OR2AG2	chr11	6768007-6768957	-	F	1151
199	HsOR11.4.2	OR2AG1	chr11	6785038-6785988	+	F	1151
200	HsOR11.4.3	OR6A2	chr11	6794725-6795708	-	F	1184
201	HsOR11.4.4	OR10A5	chr11	6845671-6846636	+	F	1166
202	HsOR11.4.5	OR10A2	chr11	6869755-6870666	+	F	1112
203	HsOR11.4.6	OR10A4	chr11	6876621-6877595	+	F	1175
204	HsOR11.4.7	OR2D2	chr11	6891574-6892500	-	F	1127

205	HsOR11.4.8	OR2D3	chr11	6921050-6921994	+	F	1145
206	HsOR11.5.5	OR5P3	chr11	7825037-7825972	-	F	1136
207	HsOR11.5.7	OR10A6	chr11	7927718-7928662	-	F	1145
208	HsOR11.5.8	OR10A3	chr11	7938576-7939520	-	F	1145
209	HsOR11.8.1	OR4B1	chr11	48216810-48217739	+	F	1130
210	HsOR11.8.3	OR4X2	chr11	48245104-48246015	+	F	1112
211	HsOR11.8.4	OR4X1	chr11	48263861-48264778	+	F	1118
212	HsOR11.8.5	OR4S1	chr11	48306223-48307152	+	F	1130
213	HsOR11.8.6	OR4C3	chr11	48325022-48325930	+	F	1109
214	HsOR11.8.13	OR4A47	chr11	48488793-48489722	+	F	1130
215	HsOR11.9.4	OR4C13	chr11	49952423-49953352	+	F	1130
216	HsOR11.9.5	OR4C12	chr11	49981572-49982501	-	F	1130
217	HsOR11.10.8	OR4C46	chr11	54603069-54603998	-	F	1130
218	HsOR11.10.6	OR4A4P	chr11	54659667-54660566	+	F	1100
219	HsOR11.10.2	OR4A5	chr11	54706885-54707832	+	F	1148
220	HsOR11.11.3	OR4A16	chr11	55343201-55344187	+	F	1187
221	HsOR11.11.4	OR4A15	chr11	55367974-55368918	+	F	1145

222	HsOR11.11.15	OR4C15	chr11	55554430-55555419	+	F	1190
223	HsOR11.11.16	OR4C16	chr11	55572128-55573060	+	F	1133
224	HsOR11.11.17	OR4C11	chr11	55603441-55604373	-	F	1133
225	HsOR11.11.18	OR4P4	chr11	55638337-55639296	+	F	1160
226	HsOR11.11.19	OR4S2	chr11	55650904-55651839	+	F	1136
227	HsOR11.11.20	OR4C6	chr11	55665167-55666096	+	F	1130
228	HsOR11.11.25	OR5D13	chr11	55773438-55774382	+	F	1145
229	HsOR11.11.27	OR5D14	chr11	55795556-55796500	+	F	1145
230	HsOR11.11.28	OR5L1	chr11	55811467-55812402	+	F	1136
231	HsOR11.11.29	OR5D18	chr11	55819630-55820571	+	F	1142
232	HsOR11.11.30	OR5L2	chr11	55827219-55828154	+	F	1136
233	HsOR11.11.31	OR5D16	chr11	55838752-55839738	+	F	1187
234	HsOR11.11.34	OR5W2	chr11	55913650-55914582	-	F	1133
235	HsOR11.11.35	OR5I1	chr11	55935456-55936400	-	F	1145
236	HsOR11.11.37	OR10AG1	chr11	55967558-55968463	-	F	1106
237	HsOR11.11.39	OR5F1	chr11	55993681-55994625	-	F	1145
238	HsOR11.11.41	OR5AS1	chr11	56030419-56031393	+	F	1175

239	HsOR11.11.45	OR8I2	chr11	56093308-56094240	+	F	1133
240	HsOR11.11.46	OR8H2	chr11	56105043-56105981	+	F	1139
241	HsOR11.11.47	OR8H3	chr11	56122373-56123311	+	F	1139
242	HsOR11.11.48	OR8J3	chr11	56136771-56137718	-	F	1148
243	HsOR11.11.49	OR8K5	chr11	56159394-56160317	-	F	1124
244	HsOR11.11.51	OR5J2	chr11	56176618-56177556	+	F	1139
245	HsOR11.11.54	OR5T2	chr11	56232106-56233095	-	F	1190
246	HsOR11.11.55	OR5T3	chr11	56252254-56253222	+	F	1169
247	HsOR11.11.56	OR5T1	chr11	56275660-56276619	+	F	1160
248	HsOR11.11.57	OR8H1	chr11	56290127-56291062	-	F	1136
249	HsOR11.11.59	OR8K3	chr11	56318307-56319245	+	F	1139
250	HsOR11.11.61	OR8K1	chr11	56346039-56346998	+	F	1160
251	HsOR11.11.62	OR8J1	chr11	56360247-56361197	+	F	1151
252	HsOR11.11.63	OR8U1	chr11	56375609-56376553	+	F	1145
253	HsOR11.11.67	OR5R1	chr11	56417258-56418232	-	F	1175
254	HsOR11.11.69	OR5M9	chr11	56462469-56463401	-	F	1133
255	HsOR11.11.70	OR5M3	chr11	56469574-56470497	-	F	1124

256	HsOR11.11.72	OR5M8	chr11	56490435-56491370	-	F	1136
257	HsOR11.11.76	OR5M11	chr11	56542340-56543257	-	F	1118
258	HsOR11.11.77	OR5M10	chr11	56576774-56577721	-	F	1148
259	HsOR11.11.79	OR5M1	chr11	56612555-56613502	-	F	1148
260	HsOR11.11.84	OR5AP2	chr11	56641489-56642442	-	F	1154
261	HsOR11.11.85	OR5AR1	chr11	56663686-56664618	+	F	1133
262	HsOR11.11.87	OR9G1	chr11	56700388-56701305	+	F	1118
263	HsOR11.11.89	OR9G4	chr11	56742828-56743766	-	F	1139
264	HsOR11.11.95	OR5AK3P	chr11	56971050-56971946	+	F	1097
265	HsOR11.11.96	OR5AK2	chr11	56988914-56989843	+	F	1130
266	HsOR11.12.3	OR6Q1	chr11	58030953-58031906	+	F	1154
267	HsOR11.12.5	OR9I1	chr11	58118500-58119444	-	F	1145
268	HsOR11.12.7	OR9Q1	chr11	58179445-58180377	+	F	1133
269	HsOR11.12.8	OR9Q2	chr11	58190491-58191435	+	F	1145
270	HsOR11.12.9	OR1S2	chr11	58203204-58204142	-	F	1139
271	HsOR11.12.10	OR1S1	chr11	58214784-58215722	+	F	1139
272	HsOR11.12.11	OR10Q1	chr11	58227916-58228875	-	F	1160

273	HsOR11.12.12	OR10W1	chr11	58266941-58267858	-	F	1118
274	HsOR11.12.17	OR5B17	chr11	58358125-58359069	-	F	1145
275	HsOR11.12.20	OR5B3	chr11	58402465-58403409	-	F	1145
276	HsOR11.12.21	OR5B2	chr11	58422332-58423261	-	F	1130
277	HsOR11.12.22	OR5B12	chr11	58439207-58440151	-	F	1145
278	HsOR11.12.23	OR5B21	chr11	58507176-58508105	-	F	1130
279	HsOR11.13.3	OR5AN1	chr11	59364459-59365394	+	F	1136
280	HsOR11.13.5	OR5A2	chr11	59421979-59422953	-	F	1175
281	HsOR11.13.6	OR5A1	chr11	59443169-59444116	+	F	1148
282	HsOR11.13.7	OR4D6	chr11	59456961-59457905	+	F	1145
283	HsOR11.13.8	OR4D10	chr11	59477430-59478365	+	F	1136
284	HsOR11.13.10	OR4D11	chr11	59503576-59504511	+	F	1136
285	HsOR11.13.11	OR4D9	chr11	59514913-59515857	+	F	1145
286	HsOR11.13.13	OR10V1	chr11	59712916-59713845	-	F	1130
287	HsOR11.16.2	OR2AT4	chr11	75088751-75089713	-	F	1163
288	HsOR11.18.1	OR6X1	chr11	123753580-123754518	-	F	1139
289	HsOR11.18.2	OR6M1	chr11	123805408-123806349	-	F	1142

290	HsOR11.18.5	OR8D4	chr11	123906432-123907376	+	F	1145
291	HsOR11.18.6	OR4D5	chr11	123939617-123940573	+	F	1157
292	HsOR11.18.7	OR6T1	chr11	123942867-123943838	-	F	1172
293	HsOR11.18.8	OR10S1	chr11	123976696-123977664	-	F	1169
294	HsOR11.18.9	OR10G6	chr11	123994163-123995161	-	F	1199
295	HsOR11.18.11	OR10G4	chr11	124015575-124016510	+	F	1136
296	HsOR11.18.12	OR10G9	chr11	124023013-124023948	+	F	1136
297	HsOR11.18.13	OR10G8	chr11	124029623-124030558	+	F	1136
298	HsOR11.18.14	OR10G7	chr11	124038066-124039001	-	F	1136
299	HsOR11.18.16	OR10D4P	chr11	124093636-124094532	-	F	1097
300	HsOR11.18.19	OR10D3	chr11	124185270-124186208	+	F	1139
301	HsOR11.18.24	OR8G1	chr11	124249676-124250611	+	F	1136
302	HsOR11.18.25	OR8G5	chr11	124264932-124265867	+	F	1136
303	HsOR11.18.26	OR8D1	chr11	124309840-124310766	-	F	1127
304	HsOR11.18.27	OR8D2	chr11	124319262-124320197	-	F	1136
305	HsOR11.18.33	OR8B2	chr11	124382402-124383343	-	F	1142
306	HsOR11.18.34	OR8B3	chr11	124396410-124397351	-	F	1142

307	HsOR11.18.35	OR8B4	chr11	124423942-124424871	-	F	1130
308	HsOR11.18.36	OR8B8	chr11	124440150-124441091	-	F	1142
309	HsOR11.18.41	OR8B12	chr11	124542722-124543654	-	F	1133
310	HsOR11.18.42	OR8A1	chr11	124570087-124571049	+	F	1163
311	HsOR12.3.1	OR10AD1	chr12	48202339-48203292	-	F	1154
312	HsOR12.3.6	OR8S1	chr12	48525632-48526570	+	F	1139
313	HsOR12.5.2	OR9K2	chr12	55129835-55130776	+	F	1142
314	HsOR12.5.5	OR10A7	chr12	55221025-55221975	+	F	1151
315	HsOR12.5.6	OR6C74	chr12	55247288-55248226	+	F	1139
316	HsOR12.5.9	OR6C6	chr12	55294288-55295232	-	F	1145
317	HsOR12.5.11	OR6C1	chr12	55320600-55321538	+	F	1139
318	HsOR12.5.12	OR6C3	chr12	55331587-55332636	+	F	1250
319	HsOR12.5.14	OR6C75	chr12	55365111-55366049	+	F	1139
320	HsOR12.5.16	OR6C66P	chr12	55388445-55389326	+	F	1082
321	HsOR12.5.17	OR6C65	chr12	55400529-55401467	+	F	1139
322	HsOR12.5.18	OR6C76	chr12	55426254-55427192	+	F	1139
323	HsOR12.5.19	OR6C2	chr12	55452214-55453152	+	F	1139

324	HsOR12.5.20	OR6C70	chr12	55469200-55470138	-	F	1139
325	HsOR12.5.21	OR6C68	chr12	55492363-55493316	+	F	1154
326	HsOR12.5.23	OR6C4	chr12	55551227-55552156	+	F	1130
327	HsOR12.5.24	OR2AP1	chr12	55574415-55575344	+	F	1130
328	HsOR12.5.26	OR10P1	chr12	55636892-55637833	+	F	1142
329	HsOR14.0.1	OR11H12	chr14	18601150-18602097	+	F	1148
330	HsOR14.1.1	OR11H2	chr14	19712936-19713883	-	F	1148
331	HsOR14.1.3	OR4Q3	chr14	19747428-19748369	+	F	1142
332	HsOR14.1.4	OR4H12P	chr14	19759910-19760788	+	F	1079
333	HsOR14.1.5	OR4M1	chr14	19780323-19781264	+	F	1142
334	HsOR14.1.7	OR4N2	chr14	19827449-19828372	+	F	1124
335	HsOR14.1.10	OR4K2	chr14	19876268-19877212	+	F	1145
336	HsOR14.1.12	OR4K5	chr14	19920607-19921578	+	F	1172
337	HsOR14.1.13	OR4K1	chr14	19935667-19936602	+	F	1136
338	HsOR14.1.15	OR4K15	chr14	19975591-19976565	+	F	1175
339	HsOR14.1.17	OR4K14	chr14	20014261-20015193	-	F	1133
340	HsOR14.1.18	OR4K13	chr14	20033844-20034758	-	F	1115

341	HsOR14.1.20	OR4L1	chr14	20060045-20060983	+	F	1139
342	HsOR14.1.22	OR4K17	chr14	20117491-20118438	+	F	1148
343	HsOR14.1.23	OR4N5	chr14	20143736-20144662	+	F	1127
344	HsOR14.1.25	OR11G2	chr14	20197438-20198373	+	F	1136
345	HsOR14.1.27	OR11H6	chr14	20223710-20224702	+	F	1193
346	HsOR14.1.29	OR11H4	chr14	20242792-20243766	+	F	1175
347	HsOR14.1.30	OR6S1	chr14	20640696-20641691	-	F	1196
348	HsOR14.2.1	OR5AU1	chr14	21154937-21155872	-	F	1136
349	HsOR14.2.2	OR10G3	chr14	21569803-21570744	-	F	1142
350	HsOR14.2.4	OR10G2	chr14	21633910-21634842	-	F	1133
351	HsOR14.2.5	OR4E2	chr14	21665083-21666024	+	F	1142
352	HsOR14.2.6	OR4E1P	chr14	21669988-21670935	-	F	1148
353	HsOR14.3.1	OR6J1	chr14	22633768-22634811	-	F	1244
354	HsOR15.1.0g	N/A	chr15	21638007-21638948	+	F	1142
355	HsOR15.1.0h	N/A	chr15	21651935-21652885	+	F	1151
356	HsOR15.1.8	OR4M2	chr15	22080625-22081566	+	F	1142
357	HsOR15.1.9	OR4N4	chr15	22094522-22095472	+	F	1151

358	HsOR15.2.1	OR4F6	chr15	101805720-101806658	+	F	1139
359	HsOR15.2.2	OR4F15	chr15	101818187-101819125	+	F	1139
360	HsOR15.2.3	OR4F14P	chr15	101828741-101829616	+	F	1076
361	HsOR15.2.6	OR4F4	chr15	101922142-101923059	-	F	1118
362	HsOR16.1.1	OR1F1	chr16	3204247-3205185	+	F	1139
363	HsOR16.1.3	OR2C1	chr16	3355941-3356879	+	F	1139
364	HsOR17.1.1	OR1D5	chr17	3062669-3063628	-	F	1160
365	HsOR17.1.2	OR1D2	chr17	3092058-3092996	-	F	1139
366	HsOR17.1.4	OR1G1	chr17	3126610-3127551	-	F	1142
367	HsOR17.1.6	OR1A2	chr17	3197519-3198448	+	F	1130
368	HsOR17.1.7	OR1A1	chr17	3215621-3216550	+	F	1130
369	HsOR17.1.10	OR3A2	chr17	3277970-3278935	-	F	1166
370	HsOR17.1.11	OR3A1	chr17	3291635-3292582	-	F	1148
371	HsOR17.1.12	OR3A4	chr17	3310311-3311357	+	F	1247
372	HsOR17.1.14	OR1E1	chr17	3397466-3398410	-	F	1145
373	HsOR17.1.15	OR3A3	chr17	3420568-3421533	+	F	1166
374	HsOR17.1.16	OR1E2	chr17	3432870-3433841	-	F	1172

375	HsOR17.2.1	OR4D1	chr17	58155154-58156086	+	F	1133
376	HsOR17.2.2	OR4D2	chr17	58169656-58170579	+	F	1124
377	HsOR19.1.3	OR4F17	chr19	110679-111596	+	F	1118
378	HsOR19.2.1	OR2Z1	chr19	8731029-8731973	+	F	1145
379	HsOR19.2.3	OR1M1	chr19	9093245-9094186	+	F	1142
380	HsOR19.2.4	OR7G2	chr19	9102269-9103243	-	F	1175
381	HsOR19.2.5	OR7G1	chr19	9114828-9115763	-	F	1136
382	HsOR19.2.7	OR7G3	chr19	9126012-9126950	-	F	1139
383	HsOR19.2.8	OR7D2	chr19	9185782-9186720	+	F	1139
384	HsOR19.2.11	OR7D4	chr19	9213899-9214837	-	F	1139
385	HsOR19.2.14	OR7E24	chr19	9251044-9252063	+	F	1220
386	HsOR19.3.1	OR7C1	chr19	14799174-14800136	-	F	1163
387	HsOR19.3.2	OR7A5	chr19	14827282-14828241	-	F	1160
388	HsOR19.3.3	OR7A10	chr19	14840948-14841877	-	F	1130
389	HsOR19.3.6	OR7A17	chr19	14880426-14881355	-	F	1130
390	HsOR19.3.11	OR7C2	chr19	14941489-14942448	+	F	1160
391	HsOR19.3.12	OR1I1	chr19	15087066-15088133	+	F	1268
392	HsOR19.4.1	OR10H2	chr19	15728044-15728991	+	F	1148

393	HsOR19.4.2	OR10H3	chr19	15741393-15742343	+	F	1151
394	HsOR19.4.3	OR10H5	chr19	15794049-15794996	+	F	1148
395	HsOR19.4.4	OR10H1	chr19	15807081-15808037	-	F	1157
396	HsOR19.4.5	OR10H4	chr19	15949008-15949958	+	F	1151
397	HsOR22.1.1	OR11H1	chr22	15528192-15529139	+	F	1148
398	HsORX.1.5	OR13H1	chrX	131544074-131545000	+	F	1127
399	Human_chr6_GL000252v2_alt_682794_683756+	OR12D1P	chr6_GL000252v2_alt	682794-683756	+	F (alt)	1163
400	Human_chr11_JH159136v1_alt_186359_187333+	OR8U8	chr11_JH159136v1_alt	186359-187333	+	F (alt)	1175
401	Human_chr11_JH159136v1_alt_193979_194908+	OR8U9	chr11_JH159136v1_alt	193979-194908	+	F (alt)	1130
402	Human_chr11_JH159137v1_alt_18234_19151+	OR9G9	chr11_JH159137v1_alt	18234-19151	+	F (alt)	1118
403	HsOR1.2.1P0	OR11I1P	chr1	110853939-110854892	+	P0	1154
404	HsOR1.4.3P0	OR10T1P	chr1	158445070-158446014	-	P0	1145
405	HsOR1.4.6P0	OR10R3P	chr1	158491219-158492160	+	P0	1142
406	HsOR1.4.19P0	OR10AA1P	chr1	158808399-158809335	-	P0	1137
407	HsOR1.4.25P0	OR10J4P	chr1	159432120-159433138	+	P0	1219
408	HsOR1.4.28P0	OR10J6P	chr1	159598298-159599227	+	P0	1130

409	HsOR1.5.11P0	OR9H1P	chr1	247774910-247775836	+	P0	1127
410	HsOR1.5.21P0	OR2L1P	chr1	247990267-247991206	+	P0	1140
411	HsOR1.5.22P0	OR2L6P	chr1	248003131-248004069	+	P0	1139
412	HsOR1.5.28P0	OR2M1P	chr1	248121932-248122882	+	P0	1151
413	HsOR3.3.1P0	OR5AC1P	chr3	98064472-98065396	+	P0	1125
414	HsOR3.3.8P	OR5H3P	chr3	98207423-98208341	+	P1	1119
415	HsOR3.3.9P	OR5H4P	chr3	98222044-98222926	+	P1	1083
416	HsOR3.3.10P0	OR5H7P	chr3	98238352-98239281	+	P0	1130
417	HsOR6.1.1P0	OR4F1P	chr6	105919-106856	-	P0	1138
418	HsOR6.2.4P0	OR2W4P	chr6	27977150-27978075	+	P0	1126
419	HsOR6.2.6P0	OR2B7P	chr6	28046434-28047415	+	P0	1182
420	HsOR6.3.7P0	OR2N1P	chr6	29137880-29138816	-	P0	1137
421	HsOR6.3.9P0	OR2J4P	chr6	29181510-29182445	+	P0	1136
422	HsOR6.3.14P0	OR2U2P	chr6	29268462-29269421	-	P0	1160
423	HsOR6.3.20P0	OR12D1P	chr6	29417280-29418226	+	P0	1147
424	HsOR6.5.1P0	OR4F7P	chr6	170639606-170640543	+	P0	1138
425	HsOR7.5.2P0	OR9A1P	chr7	141887148-141888092	+	P0	1145
426	HsOR7.6.5P0	OR10AC1P	chr7	143510933-143511911	-	P0	1179

427	HsOR7.6.8P0	OR2Q1P	chr7	143980905-143981835	+	P0	1131
428	HsOR7.6.16P0	OR2A13P	chr7	144142009-144142941	+	P0	1133
429	HsOR7.6.17P0	OR2A3P	chr7	144157226-144158159	+	P0	1134
430	HsOR7.6.20P0	OR2A20P	chr7	144250671-144251603	-	P0	1133
431	HsOR7.6.22P0	OR2A9P	chr7	144299373-144300305	+	P0	1133
432	HsOR9.1.1P0	OR13E1P	chr9	35859004-35859980	-	P0	1177
433	HsOR9.4.5P0	OR13D2P	chr9	104590017-104590953	+	P0	1137
434	HsOR9.6.7P0	OR1H1P	chr9	122607864-122608800	+	P0	1137
435	HsOR9.7.1P0	N/A	chr9	138319898-138320835	+	P0	1138
436	HsOR11.1.1P0	OR4F2P	chr11	86649-87586	-	P0	1138
437	HsOR11.3.7P0	OR52K3P	chr11	4474809-4475758	+	P0	1150
438	HsOR11.3.9P0	OR52M2P	chr11	4514942-4515924	-	P0	1183
439	HsOR11.3.15P0	OR51A9P	chr11	4660829-4661886	-	P0	1258
440	HsOR11.3.21P0	OR51F4P	chr11	4752093-4752977	-	P0	1085
441	HsOR11.3.39P0	OR51A5P	chr11	4972943-4973917	-	P0	1175
442	HsOR11.3.41P0	OR51P1P	chr11	5015138-5016079	+	P0	1142
443	HsOR11.3.46P0	OR52S1P	chr11	5076091-5077066	-	P0	1176
444	HsOR11.3.47P0	OR52E3P	chr11	5092676-5093611	+	P0	1136

445	HsOR11.3.53P	OR52Z1P	chr11	5177714-5178610	-	PF	1097
446	HsOR11.3.64P0	OR51K1P	chr11	5430653-5431602	-	P0	1150
447	HsOR11.3.71P0	OR52H2P	chr11	5551662-5552737	-	P0	1276
448	HsOR11.5.2P0	OR5P4P	chr11	7745990-7746912	+	P0	1123
449	HsOR11.5.6P0	OR5E1P	chr11	7848700-7849635	+	P0	1136
450	HsOR11.8.18P0	OR4A45P	chr11	48579436-48580354	+	P0	1119
451	HsOR11.9.8P0	N/A	chr11	50053410-50054402	-	P0	1193
452	HsOR11.10.9P0	OR4C50P	chr11	54591480-54592409	+	P0	1130
453	HsOR11.10.4P0	OR4A8P	chr11	54682877-54683824	+	P0	1148
454	HsOR11.11.9P0	OR4A13P	chr11	55466771-55467718	+	P0	1148
455	HsOR11.11.13P0	OR4C1P	chr11	55509729-55510665	+	P0	1137
456	HsOR11.11.21P0	OR4P1P	chr11	55683292-55684272	+	P0	1181
457	HsOR11.11.26P0	OR5D15P	chr11	55786965-55787909	+	P0	1145
458	HsOR11.11.32P0	OR9M1P	chr11	55855511-55856524	+	P0	1214
459	HsOR11.11.40P0	OR5F2P	chr11	56015014-56015957	-	P0	1144
460	HsOR11.11.43P0	OR5J1P	chr11	56071030-56072049	+	P0	1220
461	HsOR11.11.44P0	OR5BE1P	chr11	56082801-56083739	+	P0	1139

462	HsOR11.11.52P0	OR8V1P	chr11	56188328-56189287	-	P0	1160
463	HsOR11.11.53P0	OR8J2P	chr11	56210873-56211838	-	P0	1166
464	HsOR11.11.66P0	OR5AL1P	chr11	56412696-56413683	+	P0	1188
465	HsOR11.11.68P0	OR5M4P	chr11	56448668-56449595	+	P0	1128
466	HsOR11.11.71P0	OR5M2P	chr11	56479488-56480426	-	P0	1139
467	HsOR11.11.75P0	OR5M5P	chr11	56526564-56527520	-	P0	1157
468	HsOR11.11.78P0	OR5M13P	chr11	56597423-56598362	-	P0	1140
469	HsOR11.11.82P0	OR5M12P	chr11	56628926-56629908	+	P0	1183
470	HsOR11.11.91P0	OR5G1P	chr11	56775262-56776206	-	P0	1145
471	HsOR11.11.94P0	OR5G3P	chr11	56819573-56820546	-	P0	1174
472	HsOR11.12.1P0	OR5BA1P	chr11	57866299-57867258	-	P0	1160
473	HsOR11.12.2P0	OR5AZ1P	chr11	57917299-57918239	-	P0	1141
474	HsOR11.12.4P0	OR9I3P	chr11	58108693-58109661	-	P0	1169
475	HsOR11.12.18P0	OR5B1P	chr11	58365827-58366865	-	P0	1239
476	HsOR11.13.4P0	OR5BB1P	chr11	59391350-59392309	-	P0	1160
477	HsOR11.16.1P0	OR2AT2P	chr11	75071127-75072091	-	P0	1165
478	HsOR11.18.15P0	OR10D5P	chr11	124054793-124055732	+	P0	1140

479	HsOR11.18.18P0	OR10D1P	chr11	124158421-124159356	+	P0	1136
480	HsOR11.18.22P	OR8G2P	chr11	124224732-124225607	+	PF	1076
481	HsOR11.18.31P	OR8B1P	chr11	124365619-124366551	-	P1	1133
482	HsOR11.18.39P0	OR8B10P	chr11	124516307-124517242	-	P0	1136
483	HsOR12.3.5P0	OR8T1P	chr12	48442030-48442947	-	P0	1118
484	HsOR12.5.1P0	OR9K1P	chr12	55115533-55116474	+	P0	1142
485	HsOR12.5.25P0	OR6U2P	chr12	55611524-55612557	+	P0	1234
486	HsOR14.1.2P0	OR11K2P	chr14	19733343-19734314	-	P0	1172
487	HsOR14.1.9P0	OR4K3P	chr14	19868197-19869173	-	P0	1177
488	HsOR14.1.28P0	OR11H7P	chr14	20229402-20230346	+	P0	1145
489	HsOR15.1.1P0	N/A	chr15	21683038-21683961	+	P0	1124
490	HsOR15.1.4P0	OR11H3P	chr15	22009547-22010529	-	P0	1183
491	HsOR15.1.10P0	OR4N3P	chr15	22125511-22126434	+	P0	1124
492	HsOR15.2.5P0	OR4F28P	chr15	101875964-101876901	+	P0	1138
493	HsOR15.2.7P0	OR4G2P	chr15	101926769-101927641	-	P0	1073
494	HsOR17.1.5P0	OR1P1P	chr17	3153890-3154882	-	P0	1193
495	HsOR17.1.8P0	OR1D4	chr17	3240622-3241614	+	P0	1193

496	HsOR17.1.13P	OR1R1P	chr17	3385930-3386874	+	PF	1145
497	HsOR18.1.2P0	OR4K8P	chr18	14613210-14614184	+	P0	1175
498	HsOR19.1.4P0	OR4F8P	chr19	156279-157216	-	P0	1138
499	HsOR19.3.4P	OR7A8P	chr19	14852032-14852944	-	P1	1113
500	HsOR19.3.8P	OR7A3P	chr19	14903292-14904173	-	PF	1082
501	HsOR21.1.1P0	OR4K11P	chr21	13544210-13545185	-	P0	1176
502	Chimpanzee_AACZ04012959.1_19347_20285+	N/A	AACZ04012959.1	19347-20285	+	F	1139
503	Chimpanzee_CM000314.3_136866467_136867417-	N/A	CM000314.3	136866467-136867417	-	F	1151
504	Chimpanzee_CM000314.3_137487138_137488118+	N/A	CM000314.3	137487138-137488118	+	F	1181
505	Chimpanzee_CM000314.3_137502179_137503120+	N/A	CM000314.3	137502179-137503120	+	F	1142
506	Chimpanzee_CM000314.3_227349837_227350781+	N/A	CM000314.3	227349837-227350781	+	F	1142
507	Chimpanzee_CM000314.3_227561770_227562708+	N/A	CM000314.3	227561770-227562708	+	F	1139
508	Chimpanzee_CM000314.3_227673179_227674117+	N/A	CM000314.3	227673179-227674117	+	F	1139
509	Chimpanzee_CM000314.3_227977906_227978844+	N/A	CM000314.3	227977906-227978844	+	F	1139

510	Chimpanzee_CM000314.3_228202378_228203343-	N/A	CM000314.3	228202378-228203343	-	F	1166
511	Chimpanzee_CM000317.3_100759011_100759940+	N/A	CM000317.3	100759011-100759940	+	F	1130
512	Chimpanzee_CM000317.3_100775700_100776632+	N/A	CM000317.3	100775700-100776632	+	F	1133
513	Chimpanzee_CM000317.3_100834734_100835675+	N/A	CM000317.3	100834734-100835675	+	F	1142
514	Chimpanzee_CM000317.3_100959994_100960923+	N/A	CM000317.3	100959994-100960923	+	F	1130
515	Chimpanzee_CM000320.3_28169903_28170841+	N/A	CM000320.3	28169903-28170841	+	F	1139
516	Chimpanzee_CM000321.4_148536981_148537934+	N/A	CM000321.4	148536981-148537934	+	F	1154
517	Chimpanzee_CM000321.4_148568169_148569104+	N/A	CM000321.4	148568169-148569104	+	F	1136
518	Chimpanzee_CM000323.3_36855708_36856664-	N/A	CM000323.3	36855708-36856664	-	F	1157
519	Chimpanzee_CM000323.3_82484049_82485011-	N/A	CM000323.3	82484049-82485011	-	F	1163
520	Chimpanzee_CM000324.3_47238566_47239504+	N/A	CM000324.3	47238566-47239504	+	F	1139
521	Chimpanzee_CM000325.3_4184284_4185228+	N/A	CM000325.3	4184284-4185228	+	F	1145

522	Chimpanzee_CM000325.3_4465184_4466188-	N/A	CM000325.3	4465184-4466188	-	F	1205
523	Chimpanzee_CM000325.3_4510488_4511438-	N/A	CM000325.3	4510488-4511438	-	F	1151
524	Chimpanzee_CM000325.3_4606157_4607098+	N/A	CM000325.3	4606157-4607098	+	F	1142
525	Chimpanzee_CM000325.3_4650397_4651350+	N/A	CM000325.3	4650397-4651350	+	F	1154
526	Chimpanzee_CM000325.3_4841638_4842564+	N/A	CM000325.3	4841638-4842564	+	F	1127
527	Chimpanzee_CM000325.3_49780141_49781061+	N/A	CM000325.3	49780141-49781061	+	F	1121
528	Chimpanzee_CM000325.3_5181386_5182339+	N/A	CM000325.3	5181386-5182339	+	F	1154
529	Chimpanzee_CM000325.3_55130694_55131608+	N/A	CM000325.3	55130694-55131608	+	F	1115
530	Chimpanzee_CM000325.3_55170935_55171867+	N/A	CM000325.3	55170935-55171867	+	F	1133
531	Chimpanzee_CM000325.3_5526517_5527518+	N/A	CM000325.3	5526517-5527518	+	F	1202
532	Chimpanzee_CM000325.3_55296784_55297725-	N/A	CM000325.3	55296784-55297725	-	F	1142
533	Chimpanzee_CM000325.3_55385232_55386104+	N/A	CM000325.3	55385232-55386104	+	F	1073

534	Chimpanzee_CM000325.3_55413990_55414940+	N/A	CM000325.3	55413990-55414940	+	F	1151
535	Chimpanzee_CM000325.3_56220290_56221228-	N/A	CM000325.3	56220290-56221228	-	F	1139
536	Chimpanzee_CM000325.3_59080166_59081119-	N/A	CM000325.3	59080166-59081119	-	F	1154
537	Chimpanzee_CM000325.3_59284125_59285045+	N/A	CM000325.3	59284125-59285045	+	F	1121
538	Chimpanzee_CM000325.3_7515438_7516427+	N/A	CM000325.3	7515438-7516427	+	F	1190
539	Chimpanzee_CM000325.3_7582010_7582999-	N/A	CM000325.3	7582010-7582999	-	F	1190
540	Chimpanzee_CM000326.3_34355408_34356346-	N/A	CM000326.3	34355408-34356346	-	F	1139
541	Chimpanzee_CM000326.3_34639609_34640550-	N/A	CM000326.3	34639609-34640550	-	F	1142
542	Chimpanzee_CM000328.3_3583176_3584120+	N/A	CM000328.3	3583176-3584120	+	F	1145
543	Chimpanzee_CM000328.3_3734472_3735413+	N/A	CM000328.3	3734472-3735413	+	F	1142
544	Chimpanzee_CM000328.3_3964394_3965356+	N/A	CM000328.3	3964394-3965356	+	F	1163
545	Chimpanzee_CM000329.3_83063011_83063946-	N/A	CM000329.3	83063011-83063946	-	F	1136

546	Chimpanzee_CM000329.3_83098947_83099885-	N/A	CM000329.3	83098947-83099885	-	F	1139
547	Chimpanzee_CM000329.3_83140450_83141388+	N/A	CM000329.3	83140450-83141388	+	F	1139
548	Chimpanzee_CM000329.3_83157864_83158775-	N/A	CM000329.3	83157864-83158775	-	F	1112
549	Chimpanzee_CM000331.3_3138371_3139309-	N/A	CM000331.3	3138371-3139309	-	F	1139
550	Chimpanzee_CM000331.3_3191480_3192511+	N/A	CM000331.3	3191480-3192511	+	F	1232
551	Chimpanzee_CM000333.3_15839912_15840841-	N/A	CM000333.3	15839912-15840841	-	F	1130
552	Chimpanzee_CM000333.3_9770421_9771359-	N/A	CM000333.3	9770421-9771359	-	F	1139
553	Chimpanzee_KV421220.1_52746_53684+	N/A	KV421220.1	52746-53684	+	F	1139
554	Chimpanzee_KV421220.1_6186_7124-	N/A	KV421220.1	6186-7124	-	F	1139

a. Gene names according to the HORDE database (https://genome.weizmann.ac.il/horde/). A gene name is assigned when the chromosomal location is the same as that in Niimura et al. (2018). N/A, not applicable.
b. GRCh38/hg38 for human genes, Pantro3.0 for chimpanzee genes
c. F, intact (putatively functional) gene; PF, possibly functional gene; P0, fresh pseudogene; P1, young pseudogene (see Materials and Methods)

Appendix III. A list of neutral references considered in this study.

Sequence Name	chromosome	Start position in hg19 chromosome	End position in hg19 chromosome	Distance to the nearest gene (cM)	Distance to the nearest gene (bp)	length (bp)
hg19_NR1	chr1	5,035,005	5,037,021	0.52	191,154	2,017
hg19_NR2	chr1	14,506,986	14,513,691	1.02	392,412	6,706
hg19_NR3	chr1	14,568,485	14,569,868	1.00	355,344	1,384
hg19_NR4	chr1	30,322,942	30,324,155	1.31	669,617	1,214
hg19_NR5	chr1	218,340,347	218,346,514	0.20	112,114	6,168
hg19_NR6	chr1	244,390,422	244,391,618	0.59	124,318	1,197
hg19_NR7	chr2	2,641,137	2,642,900	0.71	306,092	1,764
hg19_NR8	chr2	2,706,019	2,709,256	1.02	370,974	3,238
hg19_NR9	chr2	2,795,128	2,796,364	1.37	396,376	1,237
hg19_NR10	chr2	5,310,848	5,311,951	1.60	520,847	1,104
hg19_NR11	chr2	6,120,819	6,122,109	0.56	279,302	1,291
hg19_NR12	chr2	6,141,407	6,142,538	0.58	299,890	1,132
hg19_NR13	chr2	6,910,442	6,912,947	0.45	67,736	2,506
hg19_NR14	chr2	6,933,376	6,934,758	0.30	45,925	1,383
hg19_NR15	chr2	129,495,232	129,496,837	1.05	419,061	1,606
hg19_NR16	chr2	235,263,600	235,264,627	0.42	137,058	1,028
hg19_NR17	chr2	240,420,252	240,428,690	0.25	97,609	8,439
hg19_NR18	chr2	240,650,039	240,652,128	0.78	244,660	2,090

hg19_NR19	chr2	240,657,232	240,658,764	0.77	238,024	1,533
hg19_NR20	chr4	1,490,613	1,494,916	0.22	100,831	4,304
hg19_NR21	chr4	162,158,922	162,160,206	0.37	144,837	1,285
hg19_NR22	chr4	165,643,860	165,648,871	0.45	226,726	5,012
hg19_NR23	chr4	180,251,672	180,256,535	2.73	1,888,015	4,864
hg19_NR24	chr4	181,310,144	181,311,461	4.42	1,933,675	1,318
hg19_NR25	chr4	181,478,614	181,481,191	4.31	1,763,945	2,578
hg19_NR26	chr4	182,186,182	182,187,330	2.61	1,057,806	1,149
hg19_NR27	chr5	2,135,181	2,136,307	0.84	252,301	1,127
hg19_NR28	chr5	2,875,151	2,877,757	0.58	119,640	2,607
hg19_NR29	chr5	26,033,083	26,034,202	1.03	846,506	1,120
hg19_NR30	chr5	105,795,249	105,796,498	0.89	916,091	1,250
hg19_NR31	chr6	791,984	796,484	0.23	135,020	4,501
hg19_NR32	chr6	23,408,367	23,409,380	1.21	717,033	1,014
hg19_NR33	chr6	92,958,053	92,963,453	0.83	986,286	5,401
hg19_NR34	chr6	169,327,939	169,329,243	0.24	259,265	1,305
hg19_NR35	chr7	41,077,035	41,078,504	0.79	176,669	1,470
hg19_NR36	chr7	42,494,455	42,497,438	0.20	217,837	2,984
hg19_NR37	chr7	79,163,487	79,165,261	0.24	80,597	1,775
hg19_NR38	chr7	148,286,192	148,287,656	0.48	ND	1,465
hg19_NR39	chr7	155,604,967	155,610,181	0.29	30,788	5,215
hg19_NR40	chr7	155,733,613	155,734,762	0.58	128,646	1,150
hg19_NR41	chr7	155,790,386	155,792,996	0.97	185,419	2,611
hg19_NR42	chr9	83,251,866	83,253,586	0.77	910,210	1,721

hg19_NR43	chr10	1,979,563	1,984,754	0.30	199,893	5,192
hg19_NR44	chr10	3,876,920	3,878,853	0.61	49,447	1,934
hg19_NR45	chr10	130,361,981	130,381,750	1.06	437,513	19,770
hg19_NR46	chr10	130,489,643	130,495,968	1.60	565,175	6,326
hg19_NR47	chr10	130,675,397	130,678,272	1.80	587,181	2,876
hg19_NR48	chr10	133,461,915	133,464,166	0.67	283,793	2,252
hg19_NR49	chr12	83,857,094	83,861,439	0.30	329,027	4,346
hg19_NR50	chr13	86,910,324	86,912,412	0.25	536,841	2,089
hg19_NR51	chr13	106,632,432	106,636,262	1.30	489,049	3,831
hg19_NR52	chr13	106,807,839	106,810,204	0.84	331,874	2,366
hg19_NR53	chr13	110,251,031	110,252,333	0.49	153,850	1,303
hg19_NR54	chr13	112,602,228	112,603,324	0.20	118,588	1,097
hg19_NR55	chr13	112,798,344	112,800,342	0.39	72,324	1,999
hg19_NR56	chr13	112,815,045	112,817,721	0.42	89,025	2,677
hg19_NR57	chr13	112,831,212	112,832,877	0.39	105,192	1,666
hg19_NR58	chr14	84,112,539	84,114,222	1.56	1,882,265	1,684
hg19_NR59	chr14	99,809,451	99,810,919	0.20	53,163	1,469
hg19_NR60	chr14	101,468,620	101,469,799	0.25	117,436	1,180
hg19_NR61	chr14	101,494,395	101,495,970	0.41	143,211	1,576
hg19_NR62	chr15	70,760,923	70,762,961	0.37	183,931	2,039
hg19_NR63	chr15	95,677,220	95,678,969	1.56	730,726	1,750
hg19_NR64	chr15	96,948,419	96,959,596	0.41	64,927	11,178
hg19_NR65	chr15	98,864,299	98,866,661	0.62	113,729	2,363
hg19_NR66	chr16	79,507,082	79,508,476	0.53	119,268	1,395

hg19_NR67	chr16	85,568,574	85,570,765	0.91	74,263	2,192
hg19_NR68	chr16	85,571,334	85,572,595	0.90	72,433	1,262
hg19_NR69	chr16	87,100,379	87,101,888	0.45	234,515	1,510
hg19_NR70	chr18	6,651,411	6,654,921	0.33	60,759	3,511
hg19_NR71	chr18	10,027,366	10,028,822	0.21	67,348	1,457
hg19_NR72	chr18	72,818,216	72,824,105	0.21	40,588	5,890
hg19_NR73	chr18	75,706,689	75,721,820	2.20	724,593	15,132
hg19_NR74	chr18	75,834,278	75,846,921	2.24	852,182	12,644
hg19_NR75	chr18	76,523,146	76,536,619	0.26	203,655	13,474
hg19_NR76	chr18	76,547,668	76,550,808	0.24	189,466	3,141
hg19_NR77	chr20	19,190,241	19,193,289	1.08	ND	3,049
hg19_NR78	chr21	26,506,065	26,507,271	0.29	450,696	1,207
hg19_NR79	chr21	28,993,938	28,995,075	1.52	654,499	1,138
hg19_NR80	chr21	40,392,775	40,393,830	0.45	153,541	1,056
hg19_NR81	chr22	44,797,744	44,799,845	0.30	88,604	2,102
hg19_NR82	chr22	49,252,397	49,253,447	0.54	104,653	1,051
hg19_NR83	chrX	9,165,357	9,166,906	0.36	164,241	1,550
hg19_NR84	chrX	124,335,707	124,337,868	0.14	116,100	2,162
hg19_NR85	chrX	124,338,204	124,340,360	0.14	113,608	2,157
			Min.	0.14	30,788	1,014
			Max.	4.42	1,933,675	19,770

Appendix IV: Mean depth of each OR gene/neutral reference region per individual				
OR gene	Mean Depth		Neutral Reference Sequence	Mean Depth
HsOR1.1.2	72.6		hg19_NR1	223.5
HsOR1.1.3	288.1		hg19_NR2	263.7
HsOR1.1.4	288.9		hg19_NR3	226.2
HsOR1.1.5	301.5		hg19_NR4	172.6
HsOR1.4.1	533.1		hg19_NR5	299.6
HsOR1.4.2	514.4		hg19_NR6	212.9
HsOR1.4.4	377.6		hg19_NR7	93.3
HsOR1.4.5	364.7		hg19_NR8	134.5
HsOR1.4.7	295.4		hg19_NR9	106.1
HsOR1.4.8	425.8		hg19_NR10	351.0
HsOR1.4.9	435.3		hg19_NR11	55.1
HsOR1.4.10	350.3		hg19_NR12	211.6
HsOR1.4.12	337.2		hg19_NR13	163.9
HsOR1.4.13	353.6		hg19_NR14	187.1
HsOR1.4.16	305.3		hg19_NR15	140.5
HsOR1.4.17	294.8		hg19_NR16	172.3
HsOR1.4.18	344.5		hg19_NR17	97.7
HsOR1.4.21	365.7		hg19_NR18	103.9
HsOR1.4.26	341.4		hg19_NR19	49.1
HsOR1.4.27	302.7		hg19_NR20	36.5
HsOR1.5.1	320.5		hg19_NR21	290.4
HsOR1.5.3	333.0		hg19_NR22	284.1
HsOR1.5.4	323.1		hg19_NR23	266.6
HsOR1.5.5	389.6		hg19_NR24	273.1
HsOR1.5.6	222.8		hg19_NR25	282.4
HsOR1.5.6b	198.9		hg19_NR26	324.5
HsOR1.5.7	224.1		hg19_NR27	90.6
HsOR1.5.8	314.3		hg19_NR28	250.1
HsOR1.5.9	373.3		hg19_NR29	314.4
HsOR1.5.10	332.5		hg19_NR30	325.5
HsOR1.5.12	260.0		hg19_NR31	232.6
HsOR1.5.14	413.9		hg19_NR32	357.1
HsOR1.5.15	344.2		hg19_NR33	283.5
HsOR1.5.16	299.2		hg19_NR34	133.8
HsOR1.5.17	414.0		hg19_NR35	214.1
HsOR1.5.18	259.3		hg19_NR36	192.9
HsOR1.5.19	199.0		hg19_NR37	360.7
HsOR1.5.23	244.6		hg19_NR38	243.7
HsOR1.5.24	357.3		hg19_NR39	141.6
HsOR1.5.25	433.1		hg19_NR40	197.4
HsOR1.5.27	322.5		hg19_NR41	83.2
HsOR1.5.29	408.7		hg19_NR42	351.6
HsOR1.5.30	422.5		hg19_NR43	233.2

HsOR1.5.31	426.8
HsOR1.5.32	286.9
HsOR1.5.33	421.0
HsOR1.5.34	413.4
HsOR1.5.35	415.3
HsOR1.5.36	341.1
HsOR1.5.37	310.5
HsOR1.5.38	286.5
HsOR1.5.39	407.8
HsOR1.5.40	322.1
HsOR1.5.41	139.9
HsOR1.5.42	108.3
HsOR1.5.43	113.2
HsOR1.5.43a	158.3
HsOR1.5.44	192.0
HsOR1.5.45	146.8
HsOR1.5.46	216.2
HsOR1.5.47	225.8
HsOR1.5.48	229.3
HsOR1.5.49	220.2
HsOR1.5.50	280.5
HsOR2.4.1	271.5
HsOR2.4.2	169.3
HsOR3.3.2	249.0
HsOR3.3.4	259.3
HsOR3.3.5	288.5
HsOR3.3.6	270.5
HsOR3.3.11	398.5
HsOR3.3.12	507.0
HsOR3.3.14	491.1
HsOR3.3.15	474.9
HsOR3.3.16	452.3
HsOR3.3.17	429.6
HsOR5.4.2	373.3
HsOR5.4.3	351.1
HsOR5.4.4	354.3
HsOR5.4.5	355.5
HsOR6.2.1	265.4
HsOR6.2.3	278.9
HsOR6.2.8	292.3
HsOR6.3.2	370.7
HsOR6.3.4	307.9
HsOR6.3.5	281.1
HsOR6.3.6	354.5
HsOR6.3.8	47.1
HsOR6.3.16	49.8

hg19_NR44	290.6
hg19_NR45	200.5
hg19_NR46	225.9
hg19_NR47	229.5
hg19_NR48	221.5
hg19_NR50	273.2
hg19_NR51	268.0
hg19_NR52	310.7
hg19_NR53	338.4
hg19_NR54	230.3
hg19_NR55	218.3
hg19_NR56	210.7
hg19_NR57	187.8
hg19_NR58	339.3
hg19_NR59	187.2
hg19_NR60	226.1
hg19_NR61	179.1
hg19_NR62	167.8
hg19_NR63	347.2
hg19_NR64	205.0
hg19_NR65	166.2
hg19_NR66	246.1
hg19_NR67	83.2
hg19_NR68	88.8
hg19_NR69	170.5
hg19_NR70	267.8
hg19_NR71	327.2
hg19_NR72	169.6
hg19_NR73	289.4
hg19_NR74	260.2
hg19_NR75	123.5
hg19_NR76	149.8
hg19_NR77	120.5
hg19_NR78	265.4
hg19_NR79	279.2
hg19_NR80	147.4
hg19_NR81	102.6
hg19_NR82	220.8
hg19_NR83	181.1
hg19_NR84	116.9
hg19_NR85	54.2

HsOR6.3.17	115.0
HsOR6.3.18	88.3
HsOR6.3.19	78.8
HsOR6.3.21	56.3
HsOR6.3.22	92.5
HsOR6.3.23	74.7
HsOR6.3.26	67.0
HsOR6.4.1	42.4
HsOR7.4.1	39.1
HsOR7.5.3	40.9
HsOR7.6.2	33.3
HsOR7.6.6	456.7
HsOR7.6.7	267.3
HsOR7.6.9	320.1
HsOR7.6.10	135.3
HsOR7.6.11	303.6
HsOR7.6.12	311.0
HsOR7.6.13	359.5
HsOR7.6.15	277.5
HsOR7.6.19	279.7
HsOR7.6.21	280.8
HsOR7.6.23	270.8
HsOR8.1.1	254.8
HsOR9.1.2	371.7
HsOR9.1.3	588.5
HsOR9.4.1	431.7
HsOR9.4.2	300.5
HsOR9.4.3	154.7
HsOR9.4.4	225.2
HsOR9.4.6	258.1
HsOR9.4.7	367.7
HsOR9.4.8	367.2
HsOR9.4.11	274.7
HsOR9.5.1	337.5
HsOR9.6.1	354.7
HsOR9.6.2	344.0
HsOR9.6.3	276.3
HsOR9.6.4	319.9
HsOR9.6.5	377.0
HsOR9.6.6	317.2
HsOR9.6.8	358.2
HsOR9.6.9	252.2
HsOR9.6.10	330.7
HsOR9.6.11	337.5
HsOR9.6.12	311.2
HsOR9.6.13	266.2

HsOR9.6.14	399.1
HsOR9.6.15	348.4
HsOR10.2.2	365.9
HsOR11.3.2	268.8
HsOR11.3.5	174.5
HsOR11.3.6	172.5
HsOR11.3.8	85.7
HsOR11.3.10	366.0
HsOR11.3.11	186.6
HsOR11.3.12	368.5
HsOR11.3.13	343.0
HsOR11.3.14	242.4
HsOR11.3.16	287.5
HsOR11.3.18	264.8
HsOR11.3.22	301.8
HsOR11.3.24	295.0
HsOR11.3.25	268.6
HsOR11.3.27	391.8
HsOR11.3.28	359.0
HsOR11.3.30	267.5
HsOR11.3.33	348.3
HsOR11.3.34	261.8
HsOR11.3.35	288.3
HsOR11.3.37	344.5
HsOR11.3.38	341.6
HsOR11.3.40	327.2
HsOR11.3.43	328.4
HsOR11.3.44	473.8
HsOR11.3.50	230.1
HsOR11.3.51	356.3
HsOR11.3.52	309.5
HsOR11.3.54	350.9
HsOR11.3.55	367.3
HsOR11.3.57	366.0
HsOR11.3.59	360.6
HsOR11.3.60	362.4
HsOR11.3.61	332.1
HsOR11.3.63	305.6
HsOR11.3.65	339.5
HsOR11.3.66	326.4
HsOR11.3.68	282.9
HsOR11.3.70	322.5
HsOR11.3.74	361.0
HsOR11.3.77	322.7
HsOR11.3.78	276.0
HsOR11.3.79	359.7

HsOR11.3.80	338.2
HsOR11.3.81	318.7
HsOR11.3.83	349.5
HsOR11.3.84	198.9
HsOR11.3.85	286.5
HsOR11.3.87	313.8
HsOR11.3.88	307.7
HsOR11.3.90	324.9
HsOR11.3.91	305.0
HsOR11.3.92	413.9
HsOR11.3.93	284.9
HsOR11.3.94	337.1
HsOR11.3.96	320.1
HsOR11.3.98	295.1
HsOR11.3.101	278.1
HsOR11.3.102	298.7
HsOR11.4.1	299.9
HsOR11.4.2	288.4
HsOR11.4.3	313.4
HsOR11.4.4	194.0
HsOR11.4.5	266.6
HsOR11.4.6	234.3
HsOR11.4.7	331.6
HsOR11.4.8	378.8
HsOR11.5.5	366.5
HsOR11.5.7	301.4
HsOR11.5.8	280.5
HsOR11.8.1	280.5
HsOR11.8.3	200.2
HsOR11.8.4	180.8
HsOR11.8.5	180.0
HsOR11.8.6	286.4
HsOR11.8.13	300.2
HsOR11.9.4	297.3
HsOR11.9.5	292.1
HsOR11.10.8	362.0
HsOR11.10.6	343.4
HsOR11.10.2	363.8
HsOR11.11.3	350.3
HsOR11.11.4	347.0
HsOR11.11.15	342.4
HsOR11.11.16	360.3
HsOR11.11.17	345.8
HsOR11.11.18	344.5
HsOR11.11.19	277.5
HsOR11.11.20	383.3

HsOR11.11.25	235.5
HsOR11.11.27	251.1
HsOR11.11.28	206.5
HsOR11.11.29	243.0
HsOR11.11.30	376.4
HsOR11.11.31	348.8
HsOR11.11.34	356.4
HsOR11.11.35	379.8
HsOR11.11.37	350.9
HsOR11.11.39	386.9
HsOR11.11.41	355.3
HsOR11.11.45	381.7
HsOR11.11.46	358.9
HsOR11.11.47	326.7
HsOR11.11.48	350.4
HsOR11.11.49	366.0
HsOR11.11.51	402.0
HsOR11.11.54	415.6
HsOR11.11.55	362.1
HsOR11.11.56	358.1
HsOR11.11.57	355.1
HsOR11.11.59	225.0
HsOR11.11.61	208.5
HsOR11.11.62	246.2
HsOR11.11.63	214.2
HsOR11.11.67	187.2
HsOR11.11.69	193.2
HsOR11.11.70	180.9
HsOR11.11.72	205.0
HsOR11.11.76	357.6
HsOR11.11.77	253.5
HsOR11.11.79	379.1
HsOR11.11.84	325.6
HsOR11.11.85	330.3
HsOR11.11.87	384.9
HsOR11.11.89	383.0
HsOR11.11.95	310.1
HsOR11.11.96	296.0
HsOR11.12.3	179.8
HsOR11.12.5	185.2
HsOR11.12.7	414.7
HsOR11.12.8	391.4
HsOR11.12.9	337.7
HsOR11.12.10	285.3
HsOR11.12.11	309.8
HsOR11.12.12	247.3

HsOR11.12.17	458.7
HsOR11.12.20	402.0
HsOR11.12.21	260.4
HsOR11.12.22	244.3
HsOR11.12.23	402.5
HsOR11.13.3	373.1
HsOR11.13.5	301.2
HsOR11.13.6	322.0
HsOR11.13.7	316.1
HsOR11.13.8	341.5
HsOR11.13.10	310.6
HsOR11.13.11	232.2
HsOR11.13.13	306.5
HsOR11.16.2	342.5
HsOR11.18.1	367.6
HsOR11.18.2	328.1
HsOR11.18.5	316.7
HsOR11.18.6	307.3
HsOR11.18.7	290.5
HsOR11.18.8	313.2
HsOR11.18.9	305.3
HsOR11.18.11	268.4
HsOR11.18.12	257.7
HsOR11.18.13	256.6
HsOR11.18.14	287.7
HsOR11.18.16	284.5
HsOR11.18.19	272.9
HsOR11.18.24	217.7
HsOR11.18.25	279.0
HsOR11.18.26	283.5
HsOR11.18.27	305.2
HsOR11.18.33	336.6
HsOR11.18.34	354.0
HsOR11.18.35	336.9
HsOR11.18.36	365.4
HsOR11.18.41	456.1
HsOR11.18.42	410.8
HsOR12.3.1	310.5
HsOR12.3.6	307.7
HsOR12.5.2	344.2
HsOR12.5.5	306.4
HsOR12.5.6	281.7
HsOR12.5.9	316.9
HsOR12.5.11	356.8
HsOR12.5.12	341.8
HsOR12.5.14	384.8

HsOR12.5.16	384.2
HsOR12.5.17	350.4
HsOR12.5.18	378.4
HsOR12.5.19	380.0
HsOR12.5.20	396.4
HsOR12.5.21	388.2
HsOR12.5.23	341.8
HsOR12.5.24	360.6
HsOR12.5.26	386.6
HsOR14.0.1	395.4
HsOR14.1.1	355.0
HsOR14.1.3	348.1
HsOR14.1.4	234.6
HsOR14.1.5	573.2
HsOR14.1.7	517.1
HsOR14.1.10	315.5
HsOR14.1.12	474.1
HsOR14.1.13	542.5
HsOR14.1.15	486.0
HsOR14.1.17	496.4
HsOR14.1.18	536.6
HsOR14.1.20	440.4
HsOR14.1.22	429.6
HsOR14.1.23	283.3
HsOR14.1.25	357.7
HsOR14.1.27	345.7
HsOR14.1.29	386.8
HsOR14.1.30	340.6
HsOR14.2.1	340.4
HsOR14.2.2	293.6
HsOR14.2.4	370.4
HsOR14.2.5	250.5
HsOR14.2.6	280.6
HsOR14.3.1	229.9
HsOR15.1.0i	282.4
HsOR15.1.0j	303.5
HsOR15.1.8	283.9
HsOR15.1.9	302.0
HsOR15.2.1	261.6
HsOR15.2.2	206.7
HsOR15.2.3	155.9
HsOR15.2.6	186.1
HsOR16.1.1	382.5
HsOR16.1.3	364.5
HsOR17.1.1	378.7
HsOR17.1.2	585.3

HsOR17.1.4	301.3
HsOR17.1.6	164.7
HsOR17.1.7	361.3
HsOR17.1.10	353.1
HsOR17.1.11	298.7
HsOR17.1.12	319.9
HsOR17.1.14	305.1
HsOR17.1.15	274.6
HsOR17.1.16	236.1
HsOR17.2.1	243.1
HsOR17.2.2	401.3
HsOR19.1.3	255.6
HsOR19.2.1	392.5
HsOR19.2.3	303.4
HsOR19.2.4	284.1
HsOR19.2.5	629.2
HsOR19.2.7	195.5
HsOR19.2.8	213.6
HsOR19.2.11	286.1
HsOR19.2.14	316.2
HsOR19.3.1	278.2
HsOR19.3.2	265.7
HsOR19.3.3	254.9
HsOR19.3.6	272.5
HsOR19.3.11	336.9
HsOR19.3.12	346.7
HsOR19.4.1	345.2
HsOR19.4.2	323.9
HsOR19.4.3	334.3
HsOR19.4.4	235.7
HsOR19.4.5	285.5
HsOR22.1.1	381.6
Human_chr6_GL000252v2_alt_682794_683756+	280.4
Human_chr11_JH159136v1_alt_186359_187333+	310.5
Human_chr11_JH159136v1_alt_193979_194908+	387.4
Human_chr11_JH159137v1_alt_18234_19151+	614.0
HsOR1.2.1P0	376.5
HsOR1.4.3P0	342.5

HsOR1.4.6P0	320.4
HsOR1.4.19P0	217.6
HsOR1.4.25P0	376.9
HsOR1.4.28P0	388.8
HsOR1.5.11P0	299.3
HsOR1.5.21P0	422.6
HsOR1.5.22P0	427.3
HsOR1.5.28P0	406.6
HsOR3.3.1P0	385.0
HsOR3.3.8P	399.9
HsOR3.3.9P	425.9
HsOR3.3.10P0	442.1
HsOR6.1.1P0	305.1
HsOR6.2.4P0	307.8
HsOR6.2.6P0	280.4
HsOR6.3.7P0	49.0
HsOR6.3.9P0	85.7
HsOR6.3.14P0	66.4
HsOR6.3.20P0	79.1
HsOR6.5.1P0	462.4
HsOR7.5.2P0	286.2
HsOR7.6.5P0	48.0
HsOR7.6.8P0	299.2
HsOR7.6.16P0	305.5
HsOR7.6.17P0	274.3
HsOR7.6.20P0	412.3
HsOR7.6.22P0	412.8
HsOR9.1.1P0	171.8
HsOR9.4.5P0	307.8
HsOR9.6.7P0	309.7
HsOR9.7.1P0	425.3
HsOR11.1.1P0	343.8
HsOR11.3.7P0	359.8
HsOR11.3.9P0	318.2
HsOR11.3.15P0	231.7
HsOR11.3.21P0	348.3
HsOR11.3.39P0	402.0
HsOR11.3.41P0	299.5
HsOR11.3.46P0	310.6
HsOR11.3.47P0	364.1
HsOR11.3.53P	292.7
HsOR11.3.64P0	293.6
HsOR11.3.71P0	382.2
HsOR11.5.2P0	366.8
HsOR11.5.6P0	173.2
HsOR11.8.18P0	238.7

HsOR11.9.8P0	361.9
HsOR11.10.9P0	358.8
HsOR11.10.4P0	350.1
HsOR11.11.9P0	341.8
HsOR11.11.13P0	297.7
HsOR11.11.21P0	279.2
HsOR11.11.26P0	341.6
HsOR11.11.32P0	368.7
HsOR11.11.40P0	340.2
HsOR11.11.43P0	340.0
HsOR11.11.44P0	339.8
HsOR11.11.52P0	379.7
HsOR11.11.53P0	198.5
HsOR11.11.66P0	347.4
HsOR11.11.68P0	376.8
HsOR11.11.71P0	322.1
HsOR11.11.75P0	294.1
HsOR11.11.78P0	300.8
HsOR11.11.82P0	279.8
HsOR11.11.91P0	181.8
HsOR11.11.94P0	162.2
HsOR11.12.1P0	263.2
HsOR11.12.2P0	341.9
HsOR11.12.4P0	273.6
HsOR11.12.18P0	409.3
HsOR11.13.4P0	343.7
HsOR11.16.1P0	312.4
HsOR11.18.15P0	314.8
HsOR11.18.18P0	279.5
HsOR11.18.22P	343.3
HsOR11.18.31P	343.8
HsOR11.18.39P0	302.3
HsOR12.3.5P0	344.3
HsOR12.5.1P0	334.2
HsOR12.5.25P0	273.3
HsOR14.1.2P0	506.6
HsOR14.1.9P0	430.0
HsOR14.1.28P0	358.3
HsOR15.1.1P0	183.4
HsOR15.1.4P0	257.2
HsOR15.1.10P0	216.3
HsOR15.2.5P0	443.1
HsOR15.2.7P0	616.8
HsOR17.1.5P0	255.6
HsOR17.1.8P0	365.0
HsOR17.1.13P	51.5

HsOR18.1.2P0	415.9		
HsOR19.1.4P0	427.5		
HsOR19.3.4P	390.9		
HsOR19.3.8P	345.6		
HsOR21.1.1P0	456.2		
HsORX.1.5	226.2		

Appendix V: Mean depth of all OR genes/neutral references averaged per individual

Sample names are given in population codes and thus described below:
HG_JPN_ANU: Ainu
AG_JPN_HNS: Honshu
HG_JPN_RYK: Ryukyu
HG_PLP_AET Aeta
HG_PLP_AGT: Agta
HG_PLP_BTK: Batak
HG_PLP_MNW: Mamanwa
AG_PLP_MNB: Manobo
AG_PLP_TLG: Tagalog
AG_PLP_VSN: Visayan
AG_EUR_DNE: Dane
AG_EUR_IRS: Irish
AG_RUS_RUS: Russian
HG_RUS_ADG: Adyghei
AG_AFR_CHG: Chagga
AG_AFR_HSA: Hausa
HG_AFR_BPG: Biaka Pygmy
HG_AFR_MPG: Mbuti Pygmy

	Autosomal genes/regions		X – chromosomal genes/regions	
Sample Name	OR genes	Neutral References	OR genes	Neutral References
HG_JPN_ANU1	664.4	465.0	382.0	198.2
HG_JPN_ANU10	496.5	359.9	516.5	276.0
HG_JPN_ANU11	565.0	416.1	579.9	324.6
HG_JPN_ANU12	591.0	429.9	341.9	188.2
HG_JPN_ANU13	553.4	406.3	321.2	169.6
HG_JPN_ANU14	532.6	390.4	315.9	166.2
HG_JPN_ANU15	545.5	396.7	619.5	314.0
HG_JPN_ANU16	562.3	404.4	613.0	323.0
HG_JPN_ANU17	521.8	376.8	510.3	300.8
HG_JPN_ANU18	372.4	270.4	391.8	210.1
HG_JPN_ANU19	502.0	361.6	512.3	277.8
HG_JPN_ANU2	555.6	373.1	387.6	190.1
HG_JPN_ANU20	728.0	509.7	781.4	400.7
HG_JPN_ANU21	414.7	307.0	442.4	237.5
HG_JPN_ANU22	623.6	457.5	374.6	194.5
HG_JPN_ANU23	575.9	410.8	336.6	165.6
HG_JPN_ANU24	720.2	521.7	435.7	223.5
HG_JPN_ANU25	439.5	320.5	246.9	128.5
HG_JPN_ANU26	605.3	434.3	631.8	340.0
HG_JPN_ANU27	590.5	425.4	619.3	325.9
HG_JPN_ANU28	534.5	384.7	575.5	289.5
HG_JPN_ANU29	414.8	279.4	439.5	214.0
HG_JPN_ANU3	491.3	360.3	512.4	265.6
HG_JPN_ANU4	527.7	381.8	319.2	172.8
HG_JPN_ANU5	539.7	389.4	578.6	309.0

HG_JPN_ANU6	495.7	350.7	512.0	277.1
HG_JPN_ANU7	612.1	434.5	357.7	183.9
HG_JPN_ANU8	587.2	415.8	580.9	326.7
HG_JPN_ANU9	538.6	392.7	315.9	160.2
AG_JPN_HNS1	553.1	369.2	319.7	164.7
AG_JPN_HNS10	474.0	352.0	255.5	147.8
AG_JPN_HNS11	377.4	270.7	200.7	113.1
AG_JPN_HNS12	410.6	285.1	221.4	112.2
AG_JPN_HNS13	508.9	357.5	289.1	144.6
AG_JPN_HNS14	288.7	197.9	153.3	78.0
AG_JPN_HNS15	580.2	404.4	316.0	168.5
AG_JPN_HNS16	801.9	574.6	472.0	246.1
AG_JPN_HNS17	598.2	379.0	323.4	159.1
AG_JPN_HNS18	418.1	334.6	225.8	127.9
AG_JPN_HNS19	241.2	204.7	122.0	79.4
AG_JPN_HNS2	393.1	264.3	236.7	118.7
AG_JPN_HNS20	532.5	374.4	301.9	166.6
AG_JPN_HNS21	290.2	199.7	164.7	80.7
AG_JPN_HNS22	487.8	345.8	264.2	149.4
AG_JPN_HNS23	264.5	188.6	147.0	77.0
AG_JPN_HNS24	585.0	393.6	327.0	169.3
AG_JPN_HNS25	311.9	213.1	161.7	85.8
AG_JPN_HNS26	528.0	364.8	323.4	158.2
AG_JPN_HNS27	267.5	188.3	151.2	75.6
AG_JPN_HNS28	509.1	353.6	279.3	144.3
AG_JPN_HNS29	510.6	362.7	280.7	148.6
AG_JPN_HNS3	471.9	336.6	267.5	138.3
AG_JPN_HNS30	602.7	411.3	360.3	175.7
AG_JPN_HNS31	298.0	221.3	143.8	89.8
AG_JPN_HNS32	609.8	441.9	331.1	187.4
AG_JPN_HNS33	877.9	607.1	493.5	261.5
AG_JPN_HNS34	569.7	403.8	305.7	173.8
AG_JPN_HNS35	381.0	281.9	216.8	117.6
AG_JPN_HNS36	753.5	527.4	414.5	222.3
AG_JPN_HNS37	361.8	251.4	203.1	101.1
AG_JPN_HNS38	285.7	200.5	142.3	82.7
AG_JPN_HNS39	328.9	228.0	172.5	86.9
AG_JPN_HNS4	476.7	330.8	256.3	140.4
AG_JPN_HNS40	543.8	350.4	306.1	152.8
AG_JPN_HNS41	502.0	354.8	498.5	267.8
AG_JPN_HNS42	608.6	418.7	330.6	168.0
AG_JPN_HNS43	430.1	302.6	226.9	130.9
AG_JPN_HNS44	528.2	353.1	303.6	146.7
AG_JPN_HNS45	470.2	329.1	254.5	132.2
AG_JPN_HNS46	426.8	304.8	251.8	119.7
AG_JPN_HNS47	524.5	371.0	274.9	155.9

AG_JPN_HNS48	537.2	359.3	308.9	150.1
AG_JPN_HNS49	571.3	395.4	334.3	162.8
AG_JPN_HNS5	369.4	258.2	200.1	106.5
AG_JPN_HNS50	489.6	327.3	258.2	142.1
AG_JPN_HNS51	583.5	413.0	326.9	175.5
AG_JPN_HNS52	495.1	351.5	270.9	139.3
AG_JPN_HNS53	521.6	371.2	293.4	152.6
AG_JPN_HNS54	589.4	419.4	343.2	172.7
AG_JPN_HNS6	51.6	31.1	27.6	12.6
AG_JPN_HNS7	603.0	398.2	354.5	162.5
AG_JPN_HNS8	310.5	210.0	177.8	87.1
AG_JPN_HNS9	530.2	366.9	286.6	151.8
HG_JPN_RYK1	168.3	89.8	81.0	40.7
HG_JPN_RYK10	176.9	92.8	98.0	40.9
HG_JPN_RYK11	181.2	115.6	91.0	48.2
HG_JPN_RYK12	65.2	34.0	30.9	16.1
HG_JPN_RYK13	173.2	97.9	93.7	40.8
HG_JPN_RYK14	198.7	105.7	99.3	47.1
HG_JPN_RYK15	81.1	50.5	41.2	19.5
HG_JPN_RYK2	177.2	96.4	95.9	42.9
HG_JPN_RYK3	143.2	79.2	73.3	34.1
HG_JPN_RYK4	127.2	65.0	69.4	28.6
HG_JPN_RYK5	242.7	127.0	121.7	57.4
HG_JPN_RYK6	181.8	96.1	95.5	41.7
HG_JPN_RYK7	203.5	108.1	105.0	47.1
HG_JPN_RYK8	159.9	85.3	78.3	38.3
HG_JPN_RYK9	136.7	71.0	64.8	32.5
HG_PLP_AET1	160.2	89.2	164.6	78.8
HG_PLP_AET10	169.1	89.7	175.2	79.2
HG_PLP_AET11	562.0	396.7	292.9	166.3
HG_PLP_AET12	166.5	90.0	173.6	79.4
HG_PLP_AET13	109.5	57.7	114.6	51.0
HG_PLP_AET14	165.8	90.2	170.3	77.1
HG_PLP_AET15	745.8	529.5	793.3	419.8
HG_PLP_AET16	88.7	49.6	97.7	40.1
HG_PLP_AET17	480.2	343.0	501.6	264.9
HG_PLP_AET18	129.6	73.7	65.6	29.7
HG_PLP_AET19	468.3	349.7	482.7	271.0
HG_PLP_AET2	455.6	342.4	477.6	263.7
HG_PLP_AET20	130.0	74.6	138.4	62.3
HG_PLP_AET21	397.7	293.7	417.9	227.0
HG_PLP_AET22	204.1	115.4	111.2	50.9
HG_PLP_AET23	161.8	92.3	81.7	39.6
HG_PLP_AET24	175.4	95.0	93.4	43.1
HG_PLP_AET25	770.0	568.0	467.0	233.8
HG_PLP_AET26	88.2	49.6	94.9	44.2

HG_PLP_AET27	543.4	383.7	320.3	158.0
HG_PLP_AET28	54.4	36.1	36.5	16.9
HG_PLP_AET29	157.6	88.6	158.6	76.5
HG_PLP_AET3	366.6	259.8	369.1	204.5
HG_PLP_AET30	505.6	340.1	296.5	152.0
HG_PLP_AET31	498.3	363.3	535.0	287.1
HG_PLP_AET32	609.2	433.8	342.6	175.2
HG_PLP_AET33	530.0	391.0	582.4	307.3
HG_PLP_AET34	501.3	360.7	288.8	148.1
HG_PLP_AET35	187.2	110.1	102.0	48.4
HG_PLP_AET36	558.5	403.5	591.0	314.9
HG_PLP_AET37	551.6	390.0	304.8	164.9
HG_PLP_AET38	488.9	336.2	285.2	152.7
HG_PLP_AET39	606.1	421.8	364.2	186.1
HG_PLP_AET4	359.6	260.6	204.7	107.7
HG_PLP_AET40	472.1	349.8	270.3	148.9
HG_PLP_AET41	560.1	402.2	379.5	200.8
HG_PLP_AET42	347.2	266.4	208.2	111.0
HG_PLP_AET43	186.7	113.8	93.4	48.8
HG_PLP_AET44	479.6	343.3	265.7	137.8
HG_PLP_AET45	438.3	310.4	254.0	132.0
HG_PLP_AET46	473.5	339.3	273.2	142.4
HG_PLP_AET47	636.8	458.8	360.7	195.1
HG_PLP_AET48	498.9	363.2	280.2	153.6
HG_PLP_AET49	499.8	354.2	269.5	141.6
HG_PLP_AET5	384.4	281.8	227.7	118.6
HG_PLP_AET50	525.7	384.7	520.3	309.5
HG_PLP_AET51	458.5	324.0	469.4	242.4
HG_PLP_AET52	465.1	333.4	488.6	267.8
HG_PLP_AET53	310.1	218.0	179.7	88.0
HG_PLP_AET54	454.1	326.1	264.6	134.2
HG_PLP_AET55	559.9	395.3	570.4	312.2
HG_PLP_AET56	449.3	319.5	485.4	245.9
HG_PLP_AET6	414.3	340.5	239.2	136.6
HG_PLP_AET7	537.6	376.5	573.3	299.9
HG_PLP_AET8	531.7	377.9	320.1	157.3
HG_PLP_AET9	197.1	108.1	107.5	48.0
HG_PLP_AGT1	498.7	366.2	286.9	149.3
HG_PLP_AGT10	436.0	313.1	457.1	240.2
HG_PLP_AGT11	510.6	342.9	514.3	266.0
HG_PLP_AGT12	508.6	360.0	569.3	292.4
HG_PLP_AGT13	552.1	388.2	561.9	300.6
HG_PLP_AGT14	560.4	395.6	589.3	317.2
HG_PLP_AGT15	630.7	427.3	670.1	340.5
HG_PLP_AGT16	551.0	380.3	632.2	292.6
HG_PLP_AGT17	480.9	344.8	495.0	272.0

HG_PLP_AGT2	61.3	38.2	38.2	19.2
HG_PLP_AGT3	689.2	460.7	399.3	195.9
HG_PLP_AGT4	79.7	49.2	52.3	21.7
HG_PLP_AGT5	605.4	431.6	354.0	178.0
HG_PLP_AGT6	692.9	513.6	380.6	218.5
HG_PLP_AGT7	615.8	399.0	361.0	169.7
HG_PLP_AGT8	75.5	50.6	54.8	20.4
HG_PLP_AGT9	340.1	244.6	360.1	192.4
HG_PLP_BTK1	979.8	711.4	576.8	323.1
HG_PLP_BTK10	208.6	119.1	109.9	51.8
HG_PLP_BTK11	180.6	104.7	100.2	47.5
HG_PLP_BTK12	68.6	42.3	81.7	36.1
HG_PLP_BTK13	93.6	65.8	131.7	52.5
HG_PLP_BTK14	60.6	35.1	38.4	15.7
HG_PLP_BTK15	182.7	105.6	106.5	45.1
HG_PLP_BTK16	598.1	434.7	351.3	186.4
HG_PLP_BTK17	127.0	83.1	160.9	73.3
HG_PLP_BTK18	574.8	412.0	608.1	321.0
HG_PLP_BTK19	469.9	330.8	540.4	255.8
HG_PLP_BTK2	122.0	74.2	75.4	31.4
HG_PLP_BTK3	81.7	63.6	57.0	25.6
HG_PLP_BTK4	410.7	308.2	227.7	124.7
HG_PLP_BTK5	609.9	420.1	337.3	186.6
HG_PLP_BTK6	359.7	261.6	190.0	108.9
HG_PLP_BTK7	523.7	384.2	306.1	165.3
HG_PLP_BTK8	103.9	66.5	66.3	31.1
HG_PLP_BTK9	165.4	105.0	92.9	43.8
HG_PLP_MNW1	117.0	62.8	122.9	55.5
HG_PLP_MNW10	278.6	147.4	284.8	132.3
HG_PLP_MNW11	144.8	76.6	147.8	69.2
HG_PLP_MNW12	209.7	118.8	108.5	53.2
HG_PLP_MNW13	188.6	102.8	94.0	50.4
HG_PLP_MNW14	135.8	73.0	74.7	32.6
HG_PLP_MNW15	413.0	298.0	241.6	125.9
HG_PLP_MNW16	571.4	396.4	340.9	169.3
HG_PLP_MNW17	71.7	41.2	37.0	17.1
HG_PLP_MNW18	512.4	346.8	295.1	147.9
HG_PLP_MNW19	587.8	422.7	341.6	182.7
HG_PLP_MNW2	59.6	31.7	26.2	13.6
HG_PLP_MNW20	568.5	406.6	594.3	316.4
HG_PLP_MNW21	92.0	52.8	52.0	23.5
HG_PLP_MNW22	60.5	33.7	66.0	30.0
HG_PLP_MNW23	117.5	65.2	60.5	29.8
HG_PLP_MNW3	140.3	75.8	77.2	32.7
HG_PLP_MNW4	173.8	96.9	95.3	44.2
HG_PLP_MNW5	520.1	369.0	286.9	159.4

HG_PLP_MNW6	579.0	399.0	614.8	321.2
HG_PLP_MNW7	88.7	53.0	93.3	43.4
HG_PLP_MNW8	449.3	327.4	493.3	258.6
HG_PLP_MNW9	154.9	81.6	157.3	77.3
AG_PLP_MNB1	127.4	72.3	133.4	61.9
AG_PLP_MNB10	177.2	95.4	98.3	41.6
AG_PLP_MNB11	205.6	112.5	226.2	100.5
AG_PLP_MNB12	197.9	109.1	105.3	49.6
AG_PLP_MNB13	158.5	85.3	179.7	79.2
AG_PLP_MNB14	28.2	16.1	27.1	13.1
AG_PLP_MNB15	434.4	311.6	452.6	245.2
AG_PLP_MNB16	592.2	412.6	629.1	328.1
AG_PLP_MNB17	544.4	395.8	570.7	312.9
AG_PLP_MNB18	539.7	380.4	550.9	307.4
AG_PLP_MNB19	603.4	432.3	626.0	342.4
AG_PLP_MNB2	93.2	52.5	98.6	45.4
AG_PLP_MNB20	563.8	414.3	316.0	175.3
AG_PLP_MNB21	572.5	413.3	337.7	170.8
AG_PLP_MNB22	513.8	345.6	560.6	285.5
AG_PLP_MNB23	556.0	389.1	583.1	314.3
AG_PLP_MNB24	602.5	438.7	326.4	180.3
AG_PLP_MNB25	535.7	388.9	294.9	154.5
AG_PLP_MNB26	591.0	384.6	347.3	176.7
AG_PLP_MNB27	731.4	543.2	750.1	417.4
AG_PLP_MNB28	533.0	370.2	555.8	306.4
AG_PLP_MNB3	124.9	70.3	62.1	29.5
AG_PLP_MNB4	177.0	102.7	184.0	88.1
AG_PLP_MNB5	136.2	76.4	67.5	34.6
AG_PLP_MNB6	161.9	93.5	170.9	77.9
AG_PLP_MNB7	205.8	110.8	218.6	99.1
AG_PLP_MNB8	171.2	99.3	170.6	86.7
AG_PLP_MNB9	48.4	26.1	50.0	21.4
AG_PLP_TLG1	208.7	126.2	231.3	105.3
AG_PLP_TLG10	167.9	98.3	185.5	89.9
AG_PLP_TLG11	148.9	82.1	77.6	38.0
AG_PLP_TLG12	109.3	56.4	53.3	27.5
AG_PLP_TLG13	31.6	16.3	16.2	8.5
AG_PLP_TLG14	47.0	24.1	44.9	24.0
AG_PLP_TLG15	50.5	29.7	51.5	26.9
AG_PLP_TLG16	561.7	386.2	570.8	304.4
AG_PLP_TLG17	595.7	407.8	644.6	328.3
AG_PLP_TLG18	484.3	358.8	501.4	295.2
AG_PLP_TLG19	619.7	414.5	633.1	327.1
AG_PLP_TLG2	295.1	165.8	308.3	143.9
AG_PLP_TLG20	572.0	419.0	337.2	170.5
AG_PLP_TLG21	576.8	404.6	586.9	314.4

AG_PLP_TLG22	561.6	415.5	311.1	182.7
AG_PLP_TLG3	48.9	28.5	24.2	11.7
AG_PLP_TLG4	112.5	65.4	57.9	29.3
AG_PLP_TLG5	76.5	43.7	76.3	38.3
AG_PLP_TLG6	23.6	12.3	11.5	6.7
AG_PLP_TLG7	86.8	50.4	46.1	21.2
AG_PLP_TLG8	117.2	70.0	64.2	32.2
AG_PLP_TLG9	54.2	31.9	24.4	13.9
AG_PLP_VSN1	425.8	298.5	439.3	255.4
AG_PLP_VSN10	567.1	421.4	333.7	180.4
AG_PLP_VSN11	364.0	269.3	193.9	106.4
AG_PLP_VSN12	303.4	224.9	158.2	86.1
AG_PLP_VSN13	637.4	448.4	363.2	195.9
AG_PLP_VSN14	556.5	393.4	564.1	314.2
AG_PLP_VSN15	478.2	350.5	252.8	146.2
AG_PLP_VSN16	622.4	464.8	648.8	380.0
AG_PLP_VSN17	575.6	381.6	281.6	170.1
AG_PLP_VSN18	703.6	524.6	413.9	222.0
AG_PLP_VSN2	551.7	396.9	575.7	307.1
AG_PLP_VSN3	539.1	417.3	316.6	176.3
AG_PLP_VSN4	442.5	328.4	272.0	141.0
AG_PLP_VSN5	491.2	363.2	469.2	290.4
AG_PLP_VSN6	569.8	411.0	324.9	168.3
AG_PLP_VSN7	581.4	429.5	614.1	346.6
AG_PLP_VSN8	517.4	375.4	518.5	282.6
AG_PLP_VSN9	493.6	377.5	287.5	161.9
AG_EUR_DNE1	88.3	55.7	41.9	21.8
AG_EUR_DNE10	107.8	67.5	57.5	26.5
AG_EUR_DNE11	92.0	57.6	51.6	24.1
AG_EUR_DNE12	97.1	56.9	99.8	47.2
AG_EUR_DNE13	112.1	73.1	59.6	28.2
AG_EUR_DNE14	200.4	118.7	100.3	51.6
AG_EUR_DNE15	100.8	67.1	49.5	26.2
AG_EUR_DNE2	51.6	32.0	27.9	13.6
AG_EUR_DNE3	63.3	38.6	33.6	16.9
AG_EUR_DNE4	69.1	43.3	31.3	17.1
AG_EUR_DNE5	127.5	74.3	67.7	30.4
AG_EUR_DNE6	143.7	83.5	72.9	37.2
AG_EUR_DNE7	130.9	80.1	69.3	31.1
AG_EUR_DNE8	70.9	43.5	34.8	18.4
AG_EUR_DNE9	98.2	64.0	52.3	24.8
AG_EUR_IRS1	115.7	67.8	54.3	27.3
AG_EUR_IRS10	97.8	51.1	55.4	22.1
AG_EUR_IRS11	227.6	119.6	112.5	55.9
AG_EUR_IRS12	187.7	96.6	104.6	43.5
AG_EUR_IRS13	149.7	85.0	74.4	34.6

AG_EUR_IRS14	155.8	82.7	85.5	39.2
AG_EUR_IRS15	476.4	327.7	234.4	135.9
AG_EUR_IRS2	68.3	38.8	38.1	17.8
AG_EUR_IRS3	112.3	63.0	53.6	28.0
AG_EUR_IRS4	132.2	82.3	67.5	33.9
AG_EUR_IRS5	124.3	64.1	68.7	30.8
AG_EUR_IRS6	183.0	96.6	109.1	41.3
AG_EUR_IRS7	164.5	85.6	149.4	62.9
AG_EUR_IRS8	145.6	83.5	75.7	35.0
AG_EUR_IRS9	93.9	49.5	44.7	21.5
AG_RUS_RUS1	236.3	125.6	129.3	56.5
AG_RUS_RUS10	95.0	52.2	50.5	22.4
AG_RUS_RUS11	99.7	55.7	51.2	24.1
AG_RUS_RUS12	131.3	68.2	62.1	32.0
AG_RUS_RUS13	149.5	83.0	77.1	35.7
AG_RUS_RUS14	100.4	53.8	54.0	23.3
AG_RUS_RUS15	154.5	85.5	79.3	36.3
AG_RUS_RUS2	63.8	38.4	34.5	15.3
AG_RUS_RUS3	132.6	72.2	64.9	30.7
AG_RUS_RUS4	146.9	75.0	78.8	34.3
AG_RUS_RUS5	178.8	95.8	90.5	43.6
AG_RUS_RUS6	101.2	58.0	52.0	23.8
AG_RUS_RUS7	153.0	84.2	76.8	37.7
AG_RUS_RUS8	162.5	88.8	85.3	38.8
AG_RUS_RUS9	117.1	60.9	61.6	28.4
HG_RUS_ADG1	148.0	93.0	81.7	36.7
HG_RUS_ADG10	258.2	153.3	133.0	66.1
HG_RUS_ADG11	144.9	87.5	77.2	36.1
HG_RUS_ADG12	245.5	152.8	129.1	62.5
HG_RUS_ADG13	59.3	38.6	28.3	14.7
HG_RUS_ADG14	230.3	144.5	119.2	62.5
HG_RUS_ADG15	231.7	148.4	122.3	59.0
HG_RUS_ADG2	172.9	106.2	173.9	86.2
HG_RUS_ADG3	110.6	65.3	118.2	49.6
HG_RUS_ADG4	141.5	88.2	70.9	34.0
HG_RUS_ADG5	90.1	54.9	79.4	36.3
HG_RUS_ADG6	84.3	49.4	49.3	22.8
HG_RUS_ADG7	124.8	75.8	134.3	63.5
HG_RUS_ADG8	88.9	51.2	97.2	43.1
HG_RUS_ADG9	75.6	45.0	77.8	36.6
AG_AFR_CHG1	148.5	88.4	80.3	37.1
AG_AFR_CHG10	90.1	54.1	46.7	21.3
AG_AFR_CHG11	152.2	90.5	77.4	38.2
AG_AFR_CHG12	53.3	29.8	27.3	13.9
AG_AFR_CHG13	83.8	54.3	44.6	19.4
AG_AFR_CHG14	147.8	86.2	81.0	37.4

AG_AFR_CHG15	110.4	68.3	60.5	27.4
AG_AFR_CHG2	145.2	87.2	77.1	35.8
AG_AFR_CHG3	77.9	47.7	81.6	36.4
AG_AFR_CHG4	258.4	158.6	142.1	62.5
AG_AFR_CHG5	86.9	53.2	40.4	21.4
AG_AFR_CHG6	78.6	44.9	41.8	19.2
AG_AFR_CHG7	106.1	64.8	59.0	26.8
AG_AFR_CHG8	148.9	93.9	79.0	36.0
AG_AFR_CHG9	101.3	60.0	49.4	25.3
AG_AFR_HSA1	87.4	55.6	43.7	20.8
AG_AFR_HSA10	103.6	57.5	66.0	24.5
AG_AFR_HSA11	157.9	94.7	83.8	40.4
AG_AFR_HSA12	79.5	48.6	42.9	19.8
AG_AFR_HSA13	68.0	36.9	62.7	28.9
AG_AFR_HSA14	39.8	22.3	40.2	19.5
AG_AFR_HSA15	107.4	65.5	56.5	27.1
AG_AFR_HSA2	48.2	30.4	23.4	11.5
AG_AFR_HSA3	104.9	65.9	57.7	28.1
AG_AFR_HSA4	287.9	182.8	141.6	69.2
AG_AFR_HSA5	127.8	75.3	67.2	30.8
AG_AFR_HSA6	105.7	66.6	50.3	26.8
AG_AFR_HSA7	66.7	40.9	34.4	16.4
AG_AFR_HSA8	78.2	47.0	75.7	33.9
AG_AFR_HSA9	76.8	37.9	44.9	16.6
HG_AFR_BPG1	109.3	65.0	55.9	27.9
HG_AFR_BPG10	278.2	164.2	150.2	71.4
HG_AFR_BPG11	59.8	38.5	30.1	15.6
HG_AFR_BPG12	99.3	62.3	54.3	23.8
HG_AFR_BPG13	168.3	105.0	88.6	41.4
HG_AFR_BPG14	135.3	83.1	70.8	33.6
HG_AFR_BPG15	138.5	85.7	81.2	34.1
HG_AFR_BPG2	136.3	87.0	65.2	35.4
HG_AFR_BPG3	92.2	52.7	51.4	23.5
HG_AFR_BPG4	116.5	70.1	60.7	30.3
HG_AFR_BPG5	28.0	20.1	14.2	8.1
HG_AFR_BPG6	108.9	66.6	58.9	27.3
HG_AFR_BPG7	130.2	78.7	65.5	32.1
HG_AFR_BPG8	136.4	83.2	71.5	34.1
HG_AFR_BPG9	61.3	38.2	33.6	15.2
HG_AFR_MPG1	101.6	64.2	52.1	25.1
HG_AFR_MPG10	121.8	75.6	125.6	57.4
HG_AFR_MPG11	135.4	86.1	77.6	32.1
HG_AFR_MPG12	125.4	76.5	62.4	29.7
HG_AFR_MPG13	159.0	88.2	79.7	35.6
HG_AFR_MPG14	81.5	48.5	40.4	20.2
HG_AFR_MPG15	67.3	44.1	36.0	16.8

HG_AFR_MPG2	108.9	71.6	55.5	26.7
HG_AFR_MPG3	61.5	38.2	30.4	15.0
HG_AFR_MPG4	52.0	32.8	56.3	24.2
HG_AFR_MPG5	64.7	37.3	33.9	15.2
HG_AFR_MPG6	109.7	66.3	50.6	25.8
HG_AFR_MPG7	96.5	51.7	48.4	22.1
HG_AFR_MPG8	91.6	54.5	45.8	22.2
HG_AFR_MPG9	122.9	69.0	66.6	29.6

Appendix VI: Disrupted Allele frequencies of OR genes found in this study

Gene	Disrupted Allele Frequency
HsOR1.4.5	1.2
HsOR1.4.7	0.1
HsOR1.4.9	55.0
HsOR1.4.10	0.4
HsOR1.4.12	0.1
HsOR1.5.4	0.4
HsOR1.5.6b	0.5
HsOR1.5.11P0	0.1
HsOR1.5.15	0.1
HsOR1.5.18	0.5
HsOR1.5.19	0.1
HsOR1.5.23	0.1
HsOR1.5.24	0.2
HsOR1.5.27	0.2
HsOR1.5.31	0.1
HsOR1.5.47	0.5
HsOR3.3.2	3.0
HsOR3.3.6	10.0
HsOR3.3.14	21.6
HsOR3.3.15	20.3
HsOR3.3.17	11.3

HsOR5.4.4	6.7
HsOR6.2.3	0.1
HsOR7.5.3	0.6
HsOR7.6.10	1.0
HsOR7.6.12	0.9
HsOR7.6.17P0	42.1
HsOR9.1.2	0.5
HsOR9.4.4	3.0
HsOR9.4.6	2.0
HsOR9.4.11	0.7
HsOR9.5.1	0.1
HsOR9.6.1	0.2
HsOR9.6.2	2.7
HsOR9.6.6	0.2
HsOR9.6.8	0.1
HsOR9.6.9	46.0
HsOR9.6.11	5.0
Chr11_JH159136v1_alt_19	1.9
HsOR11.3.2	38.5
HsOR11.3.11	1.4
HsOR11.3.22	9.5
HsOR11.3.30	0.5
HsOR11.3.35	0.5
HsOR11.3.50	0.1

HsOR11.3.51	0.1
HsOR11.3.54	1.5
HsOR11.3.57	0.1
HsOR11.3.59	1.9
HsOR11.3.63	64.5
HsOR11.3.65	26.3
HsOR11.3.66	9.2
HsOR11.3.68	0.4
HsOR11.3.70	2.4
HsOR11.3.77	0.1
HsOR11.3.78	21.6
HsOR11.3.81	1.1
HsOR11.3.87	0.1
HsOR11.3.90	0.9
HsOR11.3.91	4.5
HsOR11.3.98	0.1
HsOR11.4.3	1.5
HsOR11.4.6	0.1
HsOR11.4.7	3.6
HsOR11.4.8	4.1
HsOR11.8.1	0.1
HsOR11.8.3	0.2
HsOR11.11.3	0.1
HsOR11.11.15	0.9

HsOR11.11.16	1.0
HsOR11.11.17	3.2
HsOR11.11.18	28.9
HsOR11.11.25	0.4
HsOR11.11.30	1.0
HsOR11.11.31	0.1
HsOR11.11.37	0.7
HsOR11.11.41	0.1
HsOR11.11.45	2.4
HsOR11.11.49	0.1
HsOR11.11.72	0.1
HsOR11.11.76	2.0
HsOR11.11.77	1.7
HsOR11.11.79	12.2
HsOR11.11.84	0.2
HsOR11.11.96	2.1
HsOR11.12.3	6.7
HsOR11.12.11	0.5
HsOR11.12.17	1.6
HsOR11.12.20	1.2
HsOR11.12.22	0.1
HsOR11.13.6	0.1
HsOR11.13.7	2.5
HsOR11.13.8	5.4

HsOR11.13.10	0.1
HsOR11.13.13	3.6
HsOR11.18.1	0.2
HsOR11.18.2	0.1
HsOR11.18.5	1.7
HsOR11.18.14	0.5
HsOR11.18.27	0.1
HsOR12.3.1	15.5
HsOR12.5.6	18.8
HsOR12.5.9	2.2
HsOR12.5.12	1.1
HsOR12.5.14	5.2
HsOR12.5.17	0.6
HsOR12.5.18	3.4
HsOR12.5.23	4.1
HsOR12.5.26	0.1
HsOR14.1.3	0.2
HsOR14.1.9P0	50.0
HsOR14.1.10	0.1
HsOR14.1.12	0.1
HsOR14.1.13	0.4
HsOR14.1.20	61.3
HsOR14.1.25	0.7
HsOR14.1.27	0.1

HsOR14.1.28P0	69.6
HsOR14.1.29	0.1
HsOR14.1.30	0.1
HsOR14.2.2	0.1
HsOR14.2.5	0.7
HsOR14.2.6	4.1
HsOR15.2.1	0.5
HsOR15.2.2	0.1
HsOR16.1.3	4.1
HsOR17.1.6	0.4
HsOR17.1.11	1.1
HsOR17.2.1	0.2
HsOR19.2.1	0.1
HsOR19.2.5	1.1
HsOR19.2.7	3.2
HsOR19.2.8	0.4
HsOR19.3.8P	33.4

Appendix VII: CNV allele frequency of found in this study

OR Gene	Deletion (%)	Duplication (%)	Triplication (%)
HsOR1.1.2	2.2		
HsOR1.1.3	2.7		
HsOR1.4.7	3.0		
HsOR1.5.1		0.1	
HsOR1.5.12		0.1	
HsOR1.5.14		0.1	
HsOR1.5.15		0.2	
HsOR1.5.16		0.2	
HsOR1.5.17		0.2	
HsOR1.5.18		0.2	
HsOR1.5.19		0.1	
HsOR1.5.21P0		0.1	
HsOR1.5.22P0		0.1	
HsOR1.5.23		0.1	
HsOR1.5.24		0.1	
HsOR1.5.25		0.1	
HsOR1.5.3		0.1	
HsOR1.5.4		0.1	
HsOR1.5.46	3.5		
HsOR1.5.47	3.5		
HsOR1.5.48	1.9		

HsOR1.5.49	0.1		
HsOR1.5.5		0.1	
HsOR1.5.6		0.1	
HsOR2.4.1		0.1	
HsOR2.4.2		0.1	
HsOR21.1.1P0	0.2	0.1	
HsOR3.3.4	0.1		
HsOR5.4.3		0.4	
HsOR5.4.4		0.4	
HsOR6.3.5	5.5	3.2	0.9
HsOR6.3.6	5.5	3.7	0.9
HsOR6.3.9P0	5.5	0.6	0.6
HsOR7.6.21	0.2		
HsOR7.6.22P0	0.2		
HsOR7.6.23	0.2		
HsOR7.6.5P0	0.4		
HsOR9.1.1P0		0.2	
HsOR9.1.2		0.2	
HsOR9.1.3		0.2	
HsOR9.4.1		0.2	
HsOR9.4.11		0.2	
HsOR9.4.2		0.2	
HsOR9.4.3		0.2	
HsOR9.4.4		0.2	

HsOR9.4.5P0		0.2	
HsOR9.4.6	0.1	0.2	
HsOR9.4.7	0.7	0.2	
HsOR9.4.8	0.1	0.2	
HsOR9.5.1		0.2	
HsOR9.6.1		0.2	
HsOR9.6.10		0.2	
HsOR9.6.11		0.2	
HsOR9.6.12		0.2	
HsOR9.6.13		0.2	
HsOR9.6.14		0.2	
HsOR9.6.15		0.2	
HsOR9.6.2		0.2	
HsOR9.6.3		0.2	
HsOR9.6.4		0.2	
HsOR9.6.5		0.2	
HsOR9.6.6		0.2	
HsOR9.6.7P0		0.2	
HsOR9.6.8		0.2	
HsOR9.6.9		0.2	
HsOR9.7.1P0		0.2	
chr11_JH159136v1_alt_186359	4.0		
chr11_JH159136v1_alt_193979	4.0		
HsOR11.10.2		0.4	

HsOR11.10.4P0		0.1	
HsOR11.10.6		0.1	
HsOR11.10.8		0.1	
HsOR11.10.9P0		0.1	
HsOR11.11.17	11.0		
HsOR11.11.18	11.0		
HsOR11.11.19	11.0		
HsOR11.11.20	0.2		
HsOR11.11.21P0	14.2	20.7	
HsOR11.11.25	0.2		
HsOR11.11.26P0	0.2		
HsOR11.11.27	0.2		
HsOR11.11.28	0.2		
HsOR11.11.29	0.2		
HsOR11.11.3	0.2		
HsOR11.11.30	0.2		
HsOR11.11.62	6.2	1.7	1.0
HsOR11.11.63	7.1	3.2	0.7
HsOR11.11.70	0.1		
HsOR11.11.87	2.7		
HsOR11.12.21	0.1		
HsOR11.12.22	0.1		
HsOR11.18.22P		0.1	
HsOR11.3.101	0.1		

HsOR11.3.39P0	1.1		
HsOR11.3.43	0.1		
HsOR11.3.63	0.2		
HsOR11.3.64P0	0.2		
HsOR11.3.65	0.2		
HsOR11.3.66	0.2		
HsOR11.3.79	7.4		
HsOR11.3.80	7.4		
HsOR11.3.85	3.2		
HsOR11.3.87		3.0	
HsOR11.3.88		3.0	
HsOR11.5.6P0	0.2		
HsOR11.8.1	0.1		
HsOR11.8.13	0.1		
HsOR11.8.18P0	20.4	36.8	3.2
HsOR11.8.3	0.1		
HsOR11.8.4	0.1		
HsOR11.8.5	0.4		
HsOR11.8.6	0.1		
HsOR11.9.4		0.1	
HsOR11.9.5		0.2	
HsOR11.9.8P0		0.2	
HsOR12.3.1		0.5	
HsOR12.3.5P0		0.5	

HsOR12.3.6		0.5	
HsOR12.5.11		0.5	
HsOR12.5.12		0.5	
HsOR12.5.14		0.5	
HsOR12.5.16		2.4	
HsOR12.5.17		0.5	
HsOR12.5.18		0.5	
HsOR12.5.19		0.5	
HsOR12.5.1P0		0.5	
HsOR12.5.2		0.5	
HsOR12.5.20		0.5	
HsOR12.5.21		0.5	
HsOR12.5.23		0.5	
HsOR12.5.24		0.5	
HsOR12.5.25P0		0.5	
HsOR12.5.26		0.5	
HsOR12.5.5		0.5	
HsOR12.5.6		0.5	
HsOR12.5.9		0.5	
HsOR14.0.1		1.0	
HsOR14.1.1		1.0	
HsOR14.1.10		1.0	
HsOR14.1.12		1.0	
HsOR14.1.13		1.0	

HsOR14.1.20		0.1	
HsOR14.1.27		0.1	
HsOR14.1.28P0		0.1	
HsOR14.1.29		0.1	
HsOR14.1.2P0	0.1	1.0	
HsOR14.1.3		1.0	
HsOR14.1.4		1.0	
HsOR14.1.5		1.0	
HsOR14.1.7		1.0	
HsOR14.1.9P0		1.0	
HsOR14.3.1		0.2	
HsOR15.1.0g	1.2		
HsOR15.1.0h	1.0		
HsOR15.1.10P0	0.2	0.1	0.1
HsOR15.1.4P0	0.7		
HsOR15.1.8	0.7		
HsOR15.1.9	0.7	0.1	0.1
HsOR15.2.1		0.1	
HsOR15.2.2		0.1	
HsOR15.2.3		0.1	
HsOR15.2.5P0	0.2		
HsOR15.2.6	2.5		
HsOR15.2.7P0	2.5		
HsOR16.1.1		0.1	

HsOR16.1.3		0.2	
HsOR17.1.13P	0.1		
HsOR17.1.8P0	1.6		
HsOR18.1.2P0	1.1		
HsOR19.1.3		0.2	
HsOR19.1.4P0		0.2	
HsOR19.2.3		0.1	
HsOR19.2.4		0.1	
HsOR19.2.5		0.1	
HsOR19.2.7		0.4	
HsOR19.4.5		0.2	