PHARMACOGENETIC PROFILE OF CYP2D6 GENE IN CROATIAN ROMA

Stojanović Marković, Anita

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF BIOLOGY

Anita Stojanović Marković

PHARMACOGENETIC PROFILE OF *CYP2D6* GENE IN CROATIAN ROMA

DOCTORAL THESIS

Zagreb, 2023

Sveučilište u Zagrebu

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

BIOLOŠKI ODSJEK

Anita Stojanović Marković

FARMAKOGENETSKI PROFIL GENA *CYP2D6* ROMA U HRVATSKOJ

DOKTORSKI RAD

Zagreb, 2023.

This doctoral dissertation was made at the Institute for Anthropological Research, Zagreb, Croatia under the supervision of Professor Marijana Peričić Salihović, PhD scientific advisor with tenure as a part of the Doctoral programme of Biology at the University of Zagreb, Faculty of Science, Department of Biology.

SUPERVISOR CURRICULUM VITAE

Professor Marijana Peričić Salihović graduated in the field of molecular biology in 1993 at the Faculty of Science, University of Zagreb. At the same University, she obtained Master's degree in the field of Biological Anthropology in 1999. In 2003, she achieved the academic degree doctor of biology – field of genetics, evolution and phylogeny. She has been studying population genetics of isolated populations since 1995. In the beginnings, she studied isolated populations of the Croatian islands, working as a collaborator on several scientific projects on the subject. Studies of Croatian island populations revealed numerous genetic characteristics, which influence the health of islands` inhabitants. At the same time, she participated in the research of genetic history of the Croatian population, conducted in collaboration with colleagues from the University of Tartu, Estonia, which resulted in numerous far-reaching papers. These studies described the uniparental heritage of the Croatian population and discovered an important role of the Eastern Balkans and its ancestral populations in the formation of today's uniparental characteristics of Europeans. From 2005 on, she has been engaged in the research of the Roma population in Croatia, which resulted with more than 20 publications. These studies researched Roma`s demographic history, health status and heritable properties of pharmacogenetic variants. Under her mentorship, 12 master students did and successfully defended their graduate theses. She also supervised two doctoral theses on topics related to the population structure-analyses and one on pharmacogenomics profile of Croatian Roma.

Professor Marijana Peričić Salihović regularly publishes joint scientific papers with her students. She published 55 papers, her H-index is 20, and she was cited 2511 times.

Today professor Marijana Peričić Salihović is a scientific advisor with tenure at the Institute for anthropological research while she also teaches at the Department of Biology, Faculty of Science, University of Zagreb at graduate (Human Genome) and postgraduate (Molecular Anthropology) level.

ACKNOWLEDGEMENTS

First of all, I would like to say thank you to my mentor prof. Marijana Peričić Salihović for taking a chance on me, and giving me an opportunity to start my scientific career. Thank you for your support, guidance and help.

Special thanks go to dr. sc. Matea Zajc Petranović who went over and above her duties to help me whenever I needed a helping hand. Thank you for all proofreading, statistical analysis, all the coffee breaks, and everything else in these past four years. It would be nigh impossible to finish this thesis without your help and advice along the way.

Another thank you goes to HECUBA team (prof. Tatjana Škarić-Jurić, dr.sc. Željka Celinšćak and Maja Šetinc – future PhD). Collaborating with you was an experience like no other, and something I will not soon forget. I am thankful for our joint collaborations, and hope there will be more to come.

Last, but not least, I would also like to thank my other colleagues from the Institute for Anthropological Research that kept me sane during the trying times of my PhD that was engulfed by a world-wide pandemic, an earthquake or two, and other Herculean trials like constant relocations, among others. Thank you Ana, Olja, Petra and Vanda.

I would also like express my thanks for a great experience I got during my academic stay in Tartu, Estonia. I was there but a few month, but learned more than some can in years in a lesser environment. You have shown me what a place of knowledge and research can and should be, and given me memories that will last a lifetime. Aitäh Monika, Lena, Kristiina, Stefania, Mathilde, Tina, Bianca, Danat, Vasilij and Ajai. You have taught me a lot, and made me feel like I was home.

Special thanks to Emma, my supercycle queen, structural biology aka dark magic expert. You were the only one that could trully understand what challenges I had to go through during my PhD. Thank you for the pep talks, support, and a shoulder to cry on. I remember it all too well.

To my friends that were with me for the ride, thank you Nika, Ana, Ines, Klaudia, Lada, Josipa, and the entirety of SFera who are to numerous to name, but you know who you are. Like SFeraKon this thesis was powered by blood, sweat and tears.

At the end, I would like to say thank you to three most important people in my life, my grandmother, mother and husband. There are no words in which I could express what your support meant to me, and that I was able to share this journey with you, so I dedicate this thesis to the three of you.

University of Zagreb
Faculty of Science
Department of Biology

Doctoral disertation

PHARMACOGENETIC PROFILE OF *CYP2D6* GENE IN CROATIAN ROMA

ANITA STOJANOVIĆ MARKOVIĆ
Institute for Anthropological Research, Gajeva 32, Zagreb

The *CYP2D6* gene encodes the homonymous enzyme responsible for the metabolism of about 25% of clinically prescribed drugs, although it accounts for only 2% of all cytochromes in the human liver. The Roma are an example of a founder population where centuries of isolation and migrations have left their mark on the gene pool. Results of this study showed the influence of demographic history on single nucleotide polymorphisms in three socio-culturally different studied Roma populations. Their South Asian origin is seen in the frequencies of polymorphic variants, as well as the increased frequency of star alleles **10* and **41*. Phenotypically, normal metabolizers are the most common among Roma (51.6-65.1%), while slow metabolizers are the least common (0-7.4%). The similarity of the Roma population to European and Asian populations is evident from the LD pattern of the *CYP2D6* gene region. The three Roma groups differ significantly ($p<0.0001$) in the distribution of the five most represented haplotypes (**1, *2, *4, *10, *41*), which confirms the need to investigate them as separate groups for more accurate prescription of drugs that are metabolized by *CYP2D6*.

(134 pages, 18 pictures, 14 tables, 167 references, original in: English)
Keywords: *CYP2D6* gene, Roma population, ADME genes, pharmacogenetics, personalized medicine, single nucleotide polymorphism, haplotype

Supervisor: prof.dr.sc. Marijana Peričić Salihović, scientific advisor with tenure
Reviewers:
assoc.prof.dr.sc. Petra Korać
assoc.prof.dr.sc. Ana Galov
prof.dr.sc. Hilada Nefić

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Doktorski rad

FARMAKOGENETSKI PROFIL GENA *CYP2D6* ROMA U HRVATSKOJ

ANITA STOJANOVIĆ MARKOVIĆ
Institut za antropologiju, Gajeva 32, Zagreb

Gen *CYP2D6* kodira za istoimeni enzim odgovaran za metabolizam oko 25% klinički propisanih lijekova iako čini samo 2% svih citokroma u jetri čovjeka. Romi su primjer populacije utemeljitelja u kojoj su višestoljetna izolacija i migracije ostavile traga u njihovih zalihi gena. Rezultati ovog istraživanja su pokazali utjecaj demografske povijesti na varijacije polimorfizma jednog nukleotida u tri socio-kulturno različite, istraživane romske populacije. Njihovo južnoazijsko porijeklo očituje se u frekvencijama polimorfnih varijanti, kao i povećanoj frekvenciji zvjezdastih alela **10* i **41*. Fenotipski gledano, kod Roma je najučestaliji normalni metabolizator (51,6-65,1%), dok je spori najmanje zastupljen (0-7,4%). Sličnost Roma europskim i azijskim populacijama vidljiva je iz obrasca LD-a genske regije *CYP2D6*. Tri romske skupine se značajno razlikuju ($p<0,0001$) u distribuciji pet najzastupljenijih haplotipova (**1, *2, *4, *10, *41*), što potvrđuje potrebu da se u istraživanjima proučavaju kao zasebne grupe radi točnijeg propisivanja lijekova koje metabolizira *CYP2D6*.

(134 stranice, 18 slika, 14 tablica, 167 literaturnih navoda, jezik izvornika: engleski)
Ključne riječi: CYP2D6, Romske populacije, geni ADME, farmakogenetika, personalizirana medicina, polimorfizam jednog nukleotida, haplotip

Mentor: prof.dr.sc. Marijana Peričić Salihović, znanstvena savjetnica u trajnom zvanju
Ocjenjivači:
izv.prof.dr.sc. Petra Korać
izv.prof.dr.sc. Ana Galov
prof.dr.sc. Hilada Nefić

CONTENTS

1. INTRODUCTION

1.1. PHARMACOGENOMICS

Pharmacogenomics is a relatively new field in which genetic and genomic diversity is of great importance. It combines pharmacology and genomics to develop safe and effective medicines designed to benefit people. Currently, drugs are prescribed according to the model – one dose fits all. Today, pharmacogenetics contributes to the development of personalized medicine, which is also called individualized and/or precision medicine. Its goal is to adapt health care to each individual patient (Abrahams, 2009).

Pharmacogenomics is based on the research of how genetic variations affect the effectiveness of drugs and studies their adverse effects, mainly by analysing genes involved in **A**bsorption, **D**istribution, **M**etabolism and **E**xcretion (ADME genes). When studying drugs, researchers focus on two main factors: how much drug is needed (pharmacokinetics) and how target cells respond to the drug (pharmacodynamics).

Pharmacokinetics includes four processes: 1) Absorption (how the drug enters the bloodstream after consumption), 2) Distribution (the direction in which the drug goes after absorption and how much of the drug reaches the target cell, since many drugs cannot cross the blood-brain barrier), 3) Metabolism (how the drug is broken down in the body), and 4) Excretion (how the drug leaves the body).

Pharmacodynamics describes what a drug does to the body. Its name comes from the Greek words "phármakon", meaning drug, and "dynamikos", meaning power. Pharmacodynamics studies the biochemical, physiological, and molecular effects of drugs that include receptor binding, post-receptor effects, and chemical interactions. Examples of these types of interactions include: (1) drugs that bind to the active site of an enzyme, (2) drugs that interact with cell surface signaling proteins to disrupt downstream signaling, and (3) drugs that act by binding to molecules such as tumor necrosis factor (TNF) (Rang, 2006). Pharmacodynamics can be affected by physiological changes due to disease, aging or the presence of other drugs. Genetic mutations, thyrotoxicosis, malnutrition, myasthenia gravis and some forms of insulin-resistant diabetes are among the disorders that affect the action of the drugs. They alter receptor binding, the level of bound proteins or reduce receptor sensitivity. Age-related changes in pharmacodynamics result from changes in receptor binding or postreceptor responses. Such interactions between drugs are due to competition for receptor binding sites or an altered postreceptor response (https://www.msdmanuals.com/professional/clinical-

pharmacology/pharmacodynamics/overview-of-pharmacodynamics, accessed on 5th October 2022).

The term pharmacogenomics is often used as a synonym for pharmacogenetics. Although both terms refer to drug response related to genetic influence, pharmacogenetics focuses on the interaction of a single gene and drug, while pharmacogenomics studies the influence of many genes on the effectiveness and side effects of a particular pharmacotherapy. Therefore, pharmacogenomics includes studies of variability at the level of pharmacokinetics as well as drug pharmacodynamics (Božina, 2017).

The origins of pharmacogenetics go back to 510 BC when Pythagoras noticed that the consumption of fava beans led to a potentially fatal reaction in some but not all individuals. There is a theory that the beginnings of modern pharmacogenetics began with Snyder's 1932 study of the phenylthiourea non-taster phenotype, which is inherited as an autosomal recessive trait (Nebert, 1997, 1999).

The term pharmacogenetics was coined in 1959 by Friedrich Vogel (Vogel, 1959). In those years, therapeutic drug monitoring was introduced for the first time. It is generally defined as the clinical practice of measuring certain parameters at regular intervals that, with proper interpretation, influence drug prescribing procedures (Touw et al., 2005). This represented the first systematic approach to individualizing patient care and is still very valuable in evaluating drug therapy, especially in psychiatry and severe infections (Berm et al., 2016; Cascorbi et al., 2013). A new era of studying the genetic basis of phenotypes associated with variations in drug response began with the discovery of the structure of DNA and subsequent advances in molecular biology. Since then, the importance of genetic variants as a significant predictor of interindividual variability in drug response has been well documented (Ma & Lu, 2011).

Today, pharmacogenetics and pharmacogenomics investigate the basic connections between genes and drugs, which include genetic and genomic information, molecular and biochemical mechanisms, and functional physiological and pharmacological context, most often through clinical studies.

Pharmacogenetics aims at individualizing medications and developing regional recommendations that take population differences into account. The drugs currently in use were developed and tested primarily on people of European descent in the USA or Europe. A

thorough understanding of interethnic genetic variation in pharmacogenes and their effects on drug response is essential for effective drug prescribing worldwide (Jing Li et al., 2011).

Due to population stratification, countries with different ethnic groups are also at clinical risk of heterogeneity in drug response (Suarez-Kurtz et al., 2014). Genetic heterogeneity in drug response may also be influenced by genetic variation between admixed populations and their ancestral populations (Magalon et al., 2008). The importance of studying ethnic differences in different responses to drugs can be seen in the terms proposed for the aforementioned studies – pharmacoanthropology and ethnopharmacology (Kalow & Bertilsson, 1994).

The potential of personalized medicine is obviously interesting and significant for the pharmaceutical industry because it can improve the development, testing and registration of drugs, shorten the time between the chemical synthesis and introduction into clinical practice, and through this all reduce the overall costs of drug development (Roses, 2000).

1.2. ADME GENES

ADME is the abbreviation for Absorption, Distribution, Metabolism of the drug inside the body and for Excretion of the drug from the body. Absorption is the process through which the drug enters the bloodstream. Distribution describes the reversible transfer of a drug from one place within the body to another. Metabolism is the conversion of a generally more lipophilic xenobiotic compound into hydrophilic metabolites that can be excreted from the body. Excretion is the irreversible loss of substances from the body.

ADME genes are a group of genes encoding phase I and II metabolic enzymes, drug transporters and modifiers. Phase I enzymes are involved in oxidative, reductive or hydrolytic reactions that usually generate functional groups such as hydroxyl, carboxyl or amino groups. Typical phase I enzymes are oxidases, hydrolases, dehydrogenases and deaminases. Cytochrome P450 (CYP) enzymes such as CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP3A5 are the most important phase I enzymes for drug metabolism (Alzahrani & Rajendran, 2020).

Phase II enzymes are mainly transferases such as: UDP-glucuronosyltransferase (UGT), glutathione transferase (GST), sulfotransferase (SULT), N-acetyl transferase (NAT) and thiopurine methyltransferase (TPMT) (Jancova et al., 2010).

Phase I enzymes introduce reactive or polar groups (e.g. hydroxyl) into substrates that are often conjugated by phase II enzymes. The end products are often inactive and water soluble, making it easier for the body to eliminate them through bile, urine or stool. Consequently, phase I and II enzymes working together usually enhance drug metabolism and removal (Alzahrani & Rajendran, 2020; Hu et al., 2014; Zanger & Schwab, 2013).

Transporters include solute carrier (SLC) transporters and ATP-binding cassette (ABC) transporters. ABC transporters, such as ABCB1, ABCC2, and ABCG2, are primarily in charge of drug excretion from the cells, whereas SLC transporters, such as SLC15A2, SLC22A1, SLC22A2, SLC22A6, SLCO1B1 and SLCO1B3, are in charge of drug uptake into the cells (Nigam, 2014).

Modifiers modulate the expression of other ADME genes or affect the biochemistry of ADME enzymes (Hu et al., 2019).

In addition to drugs, ADME genes play an important role in the metabolism, transport and detoxification of a wide range of endobiotics (e.g. steroid hormones, amino acids, fatty acids, bile acids, lactate) and xenobiotics (e.g. food ingredients, pollutants and carcinogens) (Gamage et al., 2006; Hu et al., 2014; Rees et al., 2009; Zanger & Schwab, 2013).

1.3. CYTOCHROME P450

Cytochromes P450 (CYP) are a large heme-containing superfamily that plays an important role in the metabolism of drugs and xenobiotics (Estabrook, 2003). Klingenberg discovered CYP in 1954 during his research on steroid hormone metabolism when he isolated a new protein from hepatocytes (Klingenberg, 1958). Estabrook, Cooper and Rosenthal described the role of CYP as a catalyst in steroid hormone synthesis and drug metabolism (Estabrook et al., 1963). Cooper and colleagues later confirmed that CYP is a key enzyme involved in the hydroxylation reactions of drugs and steroids (Cooper et al., 1965). The wavelength at which the maximum in the absorption spectrum occurs when CYPs bind carbon monoxide in its reduced form is what gives CYP enzymes their name (Omura & Sato, 1964). The P450 superfamily originated in prokaryotes, but is now present in a wide range of organisms, including plants, animals, bacteria, viruses and humans. However, the P450 superfamily has never been shown to exist in Escherichia coli (Danielson, 2002; Urban et al., 2018).

The P450 superfamily is classified and named according to nomenclature committee guidelines based on amino acid sequence identity, phylogenetic relatedness and gene organization.

The root for all genomic and cDNA sequences is written in italics – *CYP*, and in mouse and Drosophila as *Cyp*. An individual family is designated by an Arabic numeral (members must share at least 40% amino acid identity), and a subfamily (members should share at least 55% amino acid identity) by a letter followed by a number to designate the gene. The first officially named cytochrome was *CYP1A1* (Nebert, 2002; Nelson, 1999).

P450s can be classified into four categories based on their redox partners. Class I P450s found in bacterial and eukaryotic mitochondrial membranes receive electrons from ferredoxin, an iron-sulphur protein, which is reduced by ferredoxin reductase.

Class II P450s are located in the membranes of the endoplasmic reticulum of eukaryotes and receive electrons directly from NADPH-dependent cytochrome P450 reductase (CPR), a diflavoprotein containing flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Class III P450s are self-sufficient so that the *P450* and *CPR* genes are fused into a single polypeptide. Class IV P450s can receive electrons directly from NADPH.

Eukaryotic CYPs range in size from 480 to 560 amino acids and can be grouped into three main categories based on their subcellular location. Most eukaryotic CYPs are located on the membrane of the endoplasmic reticulum (microsomal type) or mitochondria (mitochondrial type). The third category is cytosolic forms (Nelson et al., 1996).

Plant cytochromes P450 participate in a wide range of biosynthetic reactions, forming various conjugates of fatty acids, plant hormones, defence compounds or medically important drugs (Bolwell et al., 1994; Durst & O'Keefe, 1995).

Prokaryotic CYPs participate in the biosynthesis of antibiotics such as erythromycin, and catalyse key reactions responsible for environmental bioremediation through the degradation of various hydrocarbons (Chen et al., 2014).

In animal cells, most CYPs are localized on the ER membrane (Neve & Ingelman-Sundberg, 2008). In mammals, CYPs are mostly expressed in the liver, but also in extrahepatic tissues such as intestines, kidneys, lungs, heart, brain and adrenal gland (Woodland et al., 2008).

CYPs can be divided into two groups based on their substrates: biosynthetic type (B-type), and detoxifying type (D-type). The D-type is responsible for the detoxification of xenobiotics in people, which includes pharmaceuticals, plant alkaloids and aromatic chemicals. The B-type is typically involved in endogenous activities, such as the biosynthesis of hormones, steroids, and cholesterol, which have physiological effects. However, only a small number of CYPs have clearly defined substrates, and even in humans, several CYPs' substrates remain unknown (Kawashima & Satta, 2014).

In vertebrates, the number of *CYP* genes per genome varies significantly among species. Humans have 115, mice 185, and zebrafish 81 *CYP* genes (https://drnelson.uthsc.edu/, accessed on 1st October 2022). Among 115 human variants of *CYPs*, 57 are functional and 58 are pseudogenes. Functional genes are organized into 18 families and 43 subfamilies (Table 1) (Hoffman & Hu, 2007; Hu et al., 2007).

The first three families (CYP 1-3) are mainly responsible for the metabolism of exogenous substances such as drugs, while CYP families with higher numbers are generally more involved in the metabolism of endogenous substances. CYP enzymes are responsible for 75-80% of all phase I dependent metabolizers and for 65-70% of the clearance of clinically prescribed drugs (Bertz & Granneman, 1997; Evans & Relling, 1999). Human *CYP* genes are

highly polymorphic and it was necessary to develop a uniform system of nomenclature. The result is a star allele (*) nomenclature that was later adapted to describe allelic variation and haplotypes in other ADME genes as well (Robert et al., 2014). It is common for the **1* allele to represent the default reference haplotype, but it is not necessarily the most common allele in each population. When a new variant is discovered, a new unique number is given (e.g. *CYP2D6*2*). For a new number to be assigned, there must be a nucleotide change that results in an amino acid substitution, or it must be shown to affect transcription, splicing, translation, etc. A new number is not given for non-functional nucleotide changes on the same chromosome, but the major star allele is designated by an A (e.g. *CYP2D6*2A*), and the rest in additional letters (e.g. *CYP2D6*2B*) (Robarge et al., 2007). In 1999, a group of experts launched the Human Cytochrome P450 Allele Nomenclature (CYP-allele) site (www.cypalleles.ki.se), in order to centralize the available information on CYP alleles and their associated effects. Since 2017, the database has been integrated into the website www.pharmvar.org, where information is also available for other important pharmacogenes (Gaedigk et al., 2021; Gaedigk, Ingelman-Sundberg, et al., 2018; Gaedigk et al., 2020).

CYP gene variants have a major impact on individual variability in drug response and therapeutic outcomes. *CYP* genotyping and phenotyping are increasingly performed in clinical practice to identify patients at risk of drug ineffectiveness or toxicity, and to tailor individual treatments.

Table 1. List of CYP families - main functions, subfamilies and number of genes in each family, table taken from (Zhao et al., 2021)

CYP family	Primary functions	Subfamilies	Genes
1	drug metabolism	3	3
2	drug/steroid metabolism	13	16
3	drug metabolism	1	4
4	arachidonic acid/fatty acid metabolism	5	12
5	thromboxane synthase	1	1
7	steroid 7α-hydroxylase	2	2
8	bile acid biosynthesis; prostacyclin synthase	2	2
11	steroid biosynthesis	2	3
17	steroid 7α-hydroxylase	1	1
19	aromatase	1	1
20	function not determined	1	1
21	steroid biosynthesis	1	1
24	vitamin D deactivation	1	1
26	retinoic acid hydroxylase	3	3
27	bile acid biosynthesis; vitamin D3 activation	3	3
39	function not determined	1	1
46	cholesterol 24-hydroxylase	1	1
51	lanosterol 14α-demethylase	1	1

1.4. THE *CYP2D6* GENE

The CYP2D subfamily in humans consists of one functional gene (*CYP2D6*) and two pseudogenes (*CYP2D7P* and *CYP2D8P*) (Nelson, 2003; Yasukochi & Satta, 2011). This gene encodes the phase I drug-metabolizing homonymous enzyme (Heim & Meyer, 1992; Kimura et al., 1989).

The CYP2D6 protein was purified in 1984 (Distlerath & Guengerich, 1984), and in 1987, the gene was mapped to chromosome 22q13 (Eichelbaum et al., 1987). Two years later, the gene was cloned and sequenced (Kimura et al., 1989) and it was discovered that the gene locus contains two additional genes – a non-functional *CYP2D7* gene and a *CYP2D8* pseudogene (Figure 1). Compared to humans, primates and rodents have more than one functional *CYP2D* gene. Five distinctive active *CYP2D6* genes were found in rabbits, 2-3 in primates, nine in mice and six in rats (Nelson et al., 2004; Yasukochi & Satta, 2011). According to the theory, the need for detoxification potential in the diet kept the *CYP2D* genes active in mice and rats, whereas the more restrictive diet that humans previously followed and their capacity to pass on the knowledge of a healthy diet to future generations caused the loss of selective pressure, which allowed the genes to remain active (Zhou, 2018).

The *CYP2D6* gene consists of nine exons with an open reading frame of 1491 base pairs, and codes for a protein of 497 amino acids (Eichelbaum et al., 1987; Gough et al., 1993; Heim & Meyer, 1992; Kimura et al., 1989). The non-functional *CYP2D7P* gene contains a T-insertion in exon 1, disrupting the reading frame, while the pseudogene *CYP2D8P* contains multiple deletions and insertions in its exons and is an example of a true pseudogene. Based on the identification of the non-functional *CYP2D7P* gene and the *2D8P* pseudogene, Kimura et al proposed that the gene duplication leads to the generation of *CYP2D6* and *CYP2D7P*, and the gene conversion subsequently affects the formation of the *CYP2D8P* gene (Kimura et al., 1989).

The *CYP2D6* gene is one of the most studied pharmacogenes. It is extremely polymorphic with more than 100 variants and sub-variants discovered. Currently, information for 170 star alleles (*) are collected on the PharmVar Consortium website (www.pharmvar.org). Alleles can be classified into groups based on enzyme function: 1) alleles that result in increased enzyme activity, 2) alleles associated with reduced enzyme activity, 3) loss of activity (null activity) alleles, and 4) alleles associated with normal activity. Variants include single nucleotide polymorphisms (SNPs) and copy number variations (CNVs), which result from

deletions and duplications of the *CYP2D6* gene. Genetic variations of the *CYP2D6* gene affect the metabolizing activity of the CYP2D6 enzyme. Its activity can vary from complete absence to increased activity and can be grouped into four different drug metabolism phenotypes and into two subclasses (Table 2). The combination of a person's two *CYP2D6* alleles determines their diplotype, and is often used synonymously with the term "genotype" when describing a person's genetic status (Kane, 2021). The combinations of the most common diplotypes translated into phenotypes are shown in Table 3.

Table 2. Drug metabolizer phenotypes

Drug metabolizer phenotype	Allele combinations
Ultra-rapid metabolizer (UM)	One allele with normal activity and one with increased activity
Normal metabolizer (NM)	Two alleles with normal activity or One allele with increased activity and one with reduced activity
Intermediate metabolizer (IM)	Two alleles with reduced activity or One allele with normal activity and one allele with null activity
Poor metabolizer (PM)	Only null activity alleles
IM through NM	One allele with normal activity and one with reduced activity or One allele with increased activity and one with null activity
PM through IM	One allele with reduced activity and one with null activity

Table 3. *CYP2D6* allele combinations that lead to expected phenotypes

Allele	*1	*2	*3	*4	*5	*6	*7	*8	*9	*10	*17	*29	*41
*1	NM	NM	IM	IM	IM	IM	IM	IM	NM	NM	NM	NM	NM
*2		NM	IM	IM	IM	IM	IM	IM	NM	NM	NM	NM	NM
*3			PM	PM	PM	PM	PM	PM	IM	IM	IM	IM	IM
*4				PM	PM	PM	PM	PM	IM	IM	IM	IM	IM
*5					PM	PM	PM	PM	IM	IM	IM	IM	IM
*6						PM	PM	PM	IM	IM	IM	IM	IM
*7							PM	PM	IM	IM	IM	IM	IM
*8								PM	IM	IM	IM	IM	IM
*9									IM	IM	IM	IM	IM
*10										IM	IM	IM	IM
*17											IM	IM	IM
*29												IM	IM
*41													IM

* NM (normal metabolizer), IM (intermediate metabolizer), PM (poor metabolizer)

The frequency of *CYP2D6* alleles varies among different world populations. Thus, for example, *CYP2D6*4* is the most common in the populations of European origin (18%). *CYP2D6*10* is mainly found in the East and South Asian populations, where its prevalence ranges from 9% to 44%. Its frequency in the African population is between 4-6%, and among Europeans it is less than 2%. The prevalence of *CYP2D6*41* among African populations is 4-11.5%, in Asian populations 2-12%, and about 9% in European populations (Gaedigk et al., 2017; Pratt et al., 2021). The frequency of specific alleles is useful for the implementation of pharmacogenetics and genetic testing, because clinical recommendations are usually based on the phenotype of the individual. Normal metabolizers make up 43-67% of the population, and intermediate metabolizers an additional 10-44%. Poor and ultra-rapid metabolizers are less common, but these people have a higher risk of side effects or treatment failure when treated with a drug that is metabolized or bioactivated by CYP2D6 (Kane, 2021).

Although the CYP2D6 enzyme accounts for only 2-4% of the total CYP content in the liver (Williams et al., 2018; Zanger & Schwab, 2013), it is involved in the metabolism of up to 25% of drugs commonly used in medicine (Ingelman-Sundberg et al., 2007). In humans, CYP2D6 can in low to moderate levels be found in kidneys (Nishimura et al., 2003), intestines (Madani et al., 1999; Nishimura et al., 2003; Prueksaritanont et al., 1995), lungs (Bernauer et al., 2006; Guidice et al., 1997), placenta (Hakkola et al., 1996) and brain (Chinta et al., 2002; Miksys et al., 2002; Siegle et al., 2001). CYP2D6 metabolizes a wide range of drugs: antiarrhythmics, antipsychotics, antidepressants, β-blockers, anticancer drugs, several

opioid analgesics including codeine and tramadol, and many others (Beoris et al., 2016; Fleeman et al., 2010; Gaedigk, 2013; Hicks et al., 2015; Stingl et al., 2012; Zanger et al., 2008). CYP2D6 enzyme-deficiency leads to a reduced efficacy of drug therapy (Dalen et al., 1997; Sindrup et al., 1993). This enzyme has a very high affinity for alkaloids (Ingelman-Sundberg, 2004), and it is believed that its evolutionary role is related to the metabolism of alkaloids in food. There is a theory that due to the limitation of food in relation to the size of the population in north-western Africa some 10,000-20,000 years ago, there was a selection pressure that favoured the survival of subjects capable of detoxifying plant toxins to a greater extent, increasing the number of plants that can provide useful food (Aklillu et al., 2003). Different environments and diets that emerged after migration out of Africa could have exerted strong selection pressure on the *CYP2D6* and other ADME genes. Subsequently, local adaptation led to a large differentiation of the ADME genes among various populations (J. Li et al., 2011). Fuselli and colleagues proposed that the current patterns of genetic diversity in the *CYP2D6* gene are the result of selective pressure imposed by new or more concentrated CYP2D6 substrates that emerged in the food production, especially at the beginning of the Neolithic transition, in the presence of poorer dietary conditions and higher disease burdens (Fuselli et al., 2010).

1.5. PROMOTER

Transcriptional regulation plays a key role in the overall control of gene expression and downstream processes in any organism. There are complex pathways and interactions of transcription factors that regulate gene expression levels in a temporal- and/or tissue-specific manner. This is achieved by multilevel interactions between proximal and distal cis-regulatory elements, such as promoters, enhancers, silencers and locus control regions that exert their effect via ubiquitous or tissue-specific transacting protein factors that bind to assigned sequences (Georgitsi et al., 2011).

A gene promoter is a region of DNA upstream of a gene where proteins such as RNA polymerase and transcription factors bind to initiate transcription (Butler & Kadonaga, 2002; Smale & Kadonaga, 2003). The proximal promoter is a region located in close proximity (-250 to +250 base pairs) to the transcription start site (TSS). It may contain several transcription factor binding sites (TFBS) and is thought to serve as a connecting element for distal regulatory elements, allowing them to interact with the main promoter (Calhoun et al., 2002) (Figure 1). Promoters can be very complex and work together with other regions of DNA known as enhancers to ensure the most efficient transcription of that gene.

Enhancers, often known as "promoters of the promoters," activate promoters at specific locations, times and levels (Ahituv, 2012). They often show modular patterns of expression, so that a number of enhancers probably influence the gene to be active only in some tissues, while it is at the same time inactive in other tissues (Pennacchio et al., 2006; Visel et al., 2009). Enhancers can be located close to or distant from the promoter, and their function is unaffected by their position or orientation with respect to the gene they regulate. Enhancers are believed to work by attracting transcription factors (TFs), and then by physically interacting with the gene promoter (Figure 1).

There are 87 loci in the promoter and enhancer regions associated with the *CYP2D6* gene expression reported in the GeneCards database (https://www.genecards.org/, accesed on 15th September 2022). The relationship of promoter/enhancer activity and variations in the *CYP2D6* with overall drug metabolism has not been studied in detail. Several regulatory variants of the *CYP2D6* gene have been studied so far, but they were almost exclusively associated with known functional (*CYP2D6*2* and *CYP2D6*10*) or non-functional (*CYP2D6*4*) alleles (Raimundo et al., 2000). SNPs from the enhancer and promoter region may be in linkage disequilibrium (LD) with star-allele defining SNPs from the *CYP2D6* gene

region, which may affect metabolism (Gong et al., 2013).

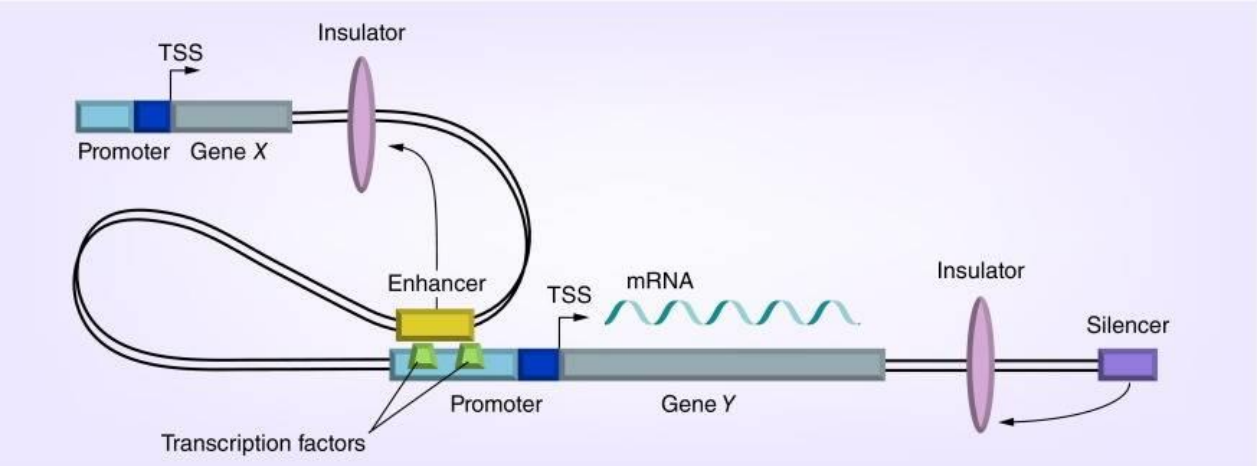


Figure 1. Regulatory elements of the gene. Promoter is shown in light blue, enhancer in yellow and transcription factors in green. Figure adapted from (Luizon & Ahituv, 2015)

1.6. ROMA POPULATION

Roma are a transnational minority population present in numerous countries of the world. They originated in India and travelled across Central Asia and modern-day Turkey in several migration waves to reach Europe in the eleventh century (Fraser, 1992). Gradual migrations over the centuries as well as socio-cultural characteristics formed the existing Roma population and left their mark in their gene pool. Roma have managed to maintain their seclusion because of their unique cultural practices and lack of contact with the general population. Different migration routes of Roma groups caused the original proto-Roma population to split into a number of smaller groups (Kalaydjieva et al., 2005). That is why the Roma groups share origins, but not recent history. Due to distinct demographic events that each group went through, such as the founder effect and the bottleneck effect, the genetic makeup of Roma groups varies from one another as do the factors that had the greatest impact (Chaix et al., 2004; Fraser, 1992; Gresham et al., 2001; Kalaydjieva, Gresham, et al., 2001; Morar et al., 2004).

Despite the Roma's lengthy history in Europe, very little was known about their origins, and what little was known was largely inaccurate. One of the misconceptions about them was that they were from Egypt, which is why the names "Gypsy" (in English) and "Gitano" (in Spanish) are used to refer to them (Fraser, 1992). It was only in the 18th century that it was assumed that they originated from India, based on linguistic research of the Romani Chib language. Samuel Augustini ab Hortis was the first to come to this conclusion in 1775 (Augustini ab Hortis, 1775). Later, Heinrich Moritz Gotlieb expanded these hypotheses, and Francz von Miklosich reconstructed the Roma's migration paths in 1873 (Miklosich, 1873).

Historical and linguistic research places the original Roma population in India (Fraser, 1992), more precisely in the provinces of Dardistan and Kafiristan (Hrvatić & Ivančić, 2000). The history of the Roma until the 14th century is mostly unclear, and the causes and beginning of migration from their ancestral homeland are still a matter of speculation, but it is assumed that the reasons for the migration were foreign military invasions and the internal social caste system (Hrvatić & Ivančić, 2000).

The immigration to Europe took place in several waves between the 10th and 16th centuries. The first groups passed through today`s Afghanistan to reach what was then Persia in the 10th century, continued to the shores of the Caspian Lake, and then split into two large

groups. The northern one moved towards Armenia (later continued towards Russia), and the southern group moved along the Euphrates and Tigris rivers towards Syria and Egypt (Fraser, 1992). A part of the southern group separated and reached Spain via Gibraltar through Africa (Hrvatić & Ivančić, 2000).

The largest part of the Roma came to the Balkans between the 11th and 12th centuries, passing through Anatolia and crossing the Bosphorus (Hrvatić & Ivančić, 2000). After arriving in the Balkans, the majority of Roma settled there to this day. Another part of the Roma continued towards the Western and Northern Europe, and the third one crossed the Danube and settled in present-day Romania, i.e. the former provinces of Wallachia, Moldavia and Transylvania, where they ended up in slavery (Marushiakova & Popov, 2001). The first wave of Roma migrations is shown in Figure 2.

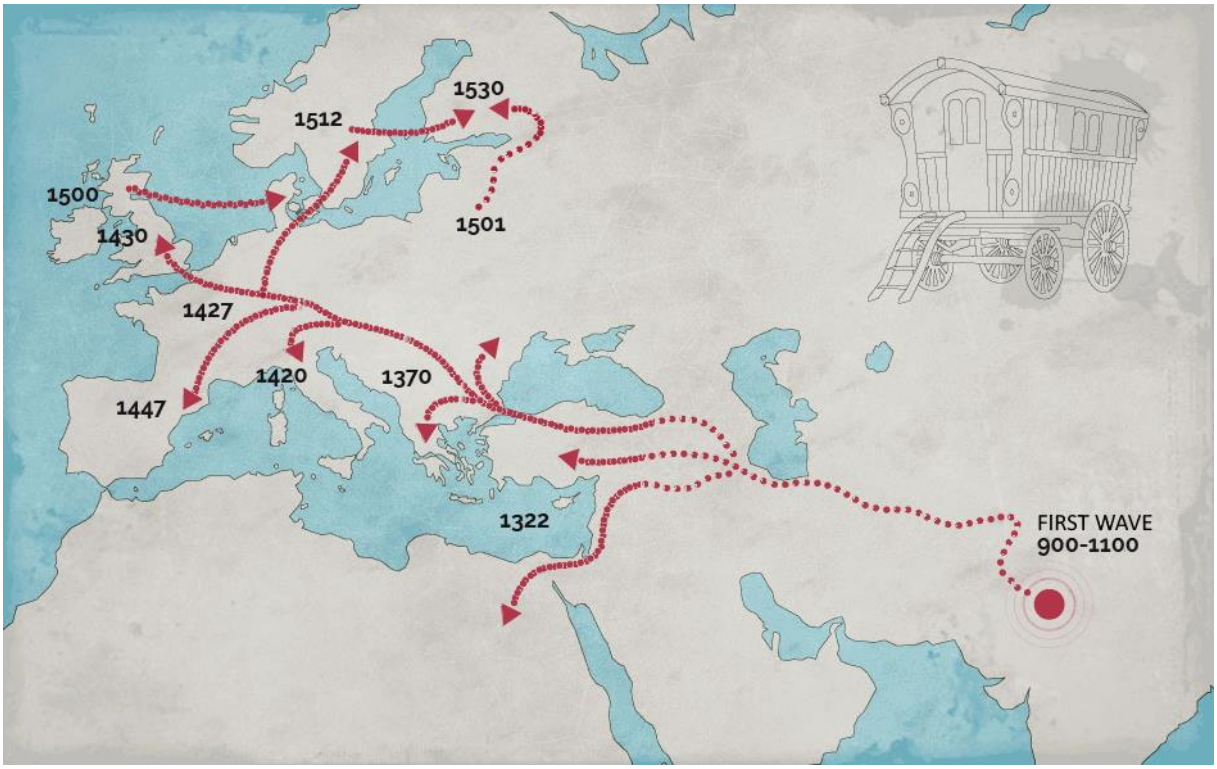


Figure 2. Roma's first wave of migration from India to Europe (adapted from roma.inantro.hr)

The slavery of this group of Roma lasted for the next 500 years, and today we call these Roma Vlax Roma or Bayash/Boyash Roma. After the abolition of slavery in 1856, these Roma emigrated to the west in several waves (Fraser, 1992). During these migrations, a part of Bayash Roma came to Croatia and mostly settled in areas along large rivers. Other Bayash Roma groups spread throughout the rest of Europe, and some migrated overseas in countries such as the USA and Canada. The second wave of migration of Roma is shown in Figure 3.

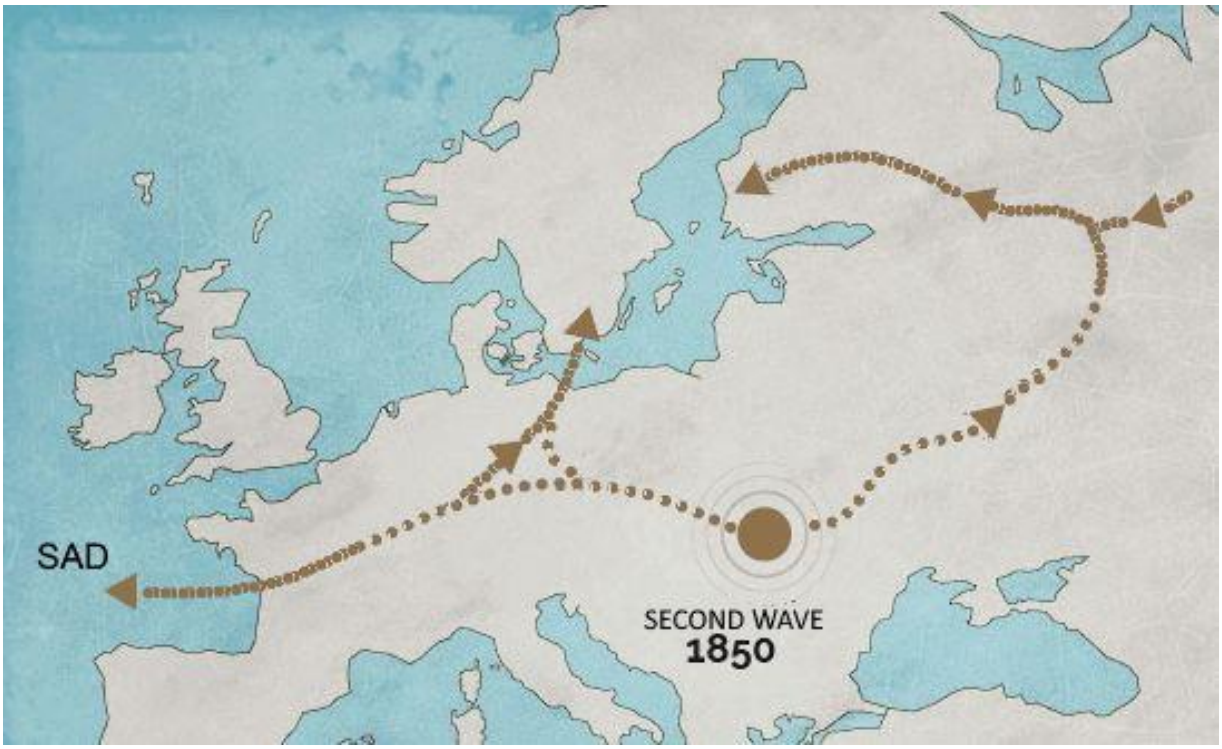


Figure 3. Roma's second emigrational wave inside Europe and towards USA (adapted from roma.inantro.hr)

The third wave of Roma migration took place throughout the 20th century, mostly for economic reasons. After the Second World War, there was a large emigration of the European population to America, and Roma were among them. A similar migration of Roma took place in the 90s of the 20th century, when many Roma moved to the countries of Western Europe and America (Marushiakova & Popov, 2001). The third wave of Roma migration is shown in Figure 4.

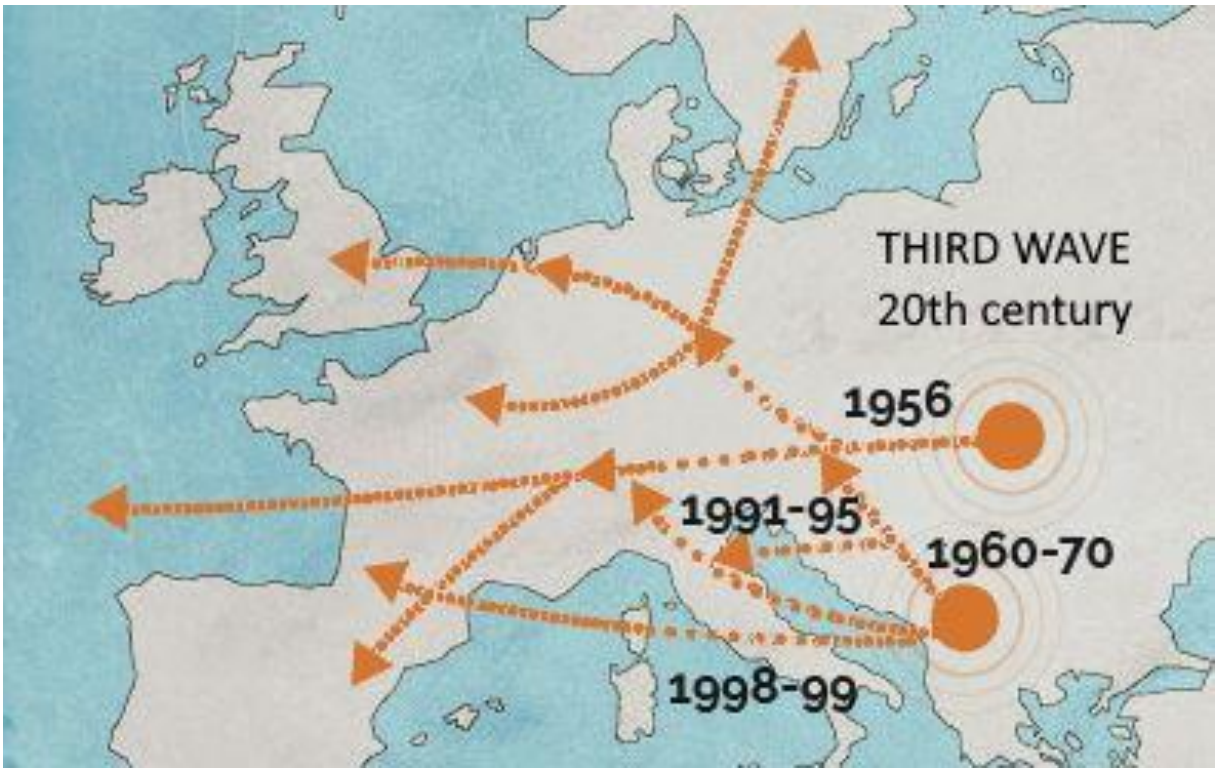


Figure 4. The third wave of Roma migration after WWII and in the 90s of the 20th century (adapted from roma.inantro.hr)

It is estimated that around 15 million Roma live in the world today, of which around 12 million live in Europe. Roma are one of the 22 national minorities recognized in Croatia. According to the official population Census from 2011, 16,975 Roma lived in Croatia, and according to the 2021 Census, there are currently 17,980 of them. Their number by county is shown in Table 4. Most Roma live in Medjimurje County, Osijek-Baranja County, Sisak-Moslavina County and in the City of Zagreb. It should be noted that Roma often do not want to declare themselves on the official list due to various fears and cautions they have developed during centuries of persecution, so the official number is not entirely reliable. It is estimated that between 30,000 and 40,000 Roma live in Croatia (Liégeois, 2009).

Table 4. Number of Roma living in Croatia according to the Censuses from 2011 and 2021 (source: Croatian Bureau of Statistics)

	2011			2021		
County	**Total**	**Roma**	**%**	**Total**	**Roma**	**%**
Zagreb County	317,606	258	0.08	299,985	245	0.08
Krapina-Zagorje	132,892	3	0.00	120,702	13	0.01
Sisak-Moslavina	172,439	1,463	0.85	139,603	1,636	1.17
Karlovac	128,899	26	0.02	112,195	55	0.05
Varazdin	175,951	711	0.40	159,487	1,101	0.69
Koprivnica-Krizevci	115,584	925	0.80	101,221	1,052	1.04
Bjelovar-Bilogora	119,764	391	0.33	101,879	381	0.37
Primorje-Gorski Kotar	296,195	1,072	0.36	265,419	683	0.26
Lika-Senj	50,927	21	0.04	42,748	28	0.07
Virovitica-Podravina	84,836	14	0.02	70,368	38	0.05
Pozega-Slavonia	78,034	13	0.02	64,084	30	0.05
Brod-Posavina	158,575	1,178	0.74	130,267	1,155	0.89
Zadar	170,017	12	0.01	159,766	16	0.01
Osijek-Baranja	305,032	1,874	0.61	258,026	1,660	0.64
Sibenik-Knin	109,375	22	0.02	96,381	12	0.01
Vukovar-Syrmia	179,521	253	0.14	143,113	191	0.13
Split-Dalmatia	454,798	8	0.00	423,407	28	0.01
Istria	208,055	858	0.41	195,237	531	0.27
Dubrovnik-Neretva	122,568	11	0.01	115,564	4	0.00
Medjimurje	113,804	5,107	4.49	105,250	6,954	6.61
City of Zagreb	790,017	2,755	0.35	767,131	2,167	0.28
Republic of Croatia	**4,284,889**	**16,975**	**0.40**	**3,871,833**	**17,980**	**0.46**

A group is the basic unit of Roma society, it is determined by customs, region, professions, own traditions, language/dialect, religion, and often has its own self-governing bodies (Fraser, 1992; Reyniers, 1995). Endogamy is the norm of many groups, which means finding a partner exclusively within the members of one's own group. The greatest diversity of groups is found in the Balkans, with more than 50 different Roma groups in Bulgaria alone (Marushiakova & Popov, 2001).

Throughout their history, Roma were mostly nomads and moved between places in caravans. The exception was the period of slavery in Romania when most of them were forced to a sedentary life because they were considered property of a monasteries or settlements, so their movement outside those places was forbidden. The Roma who belonged to the rulers were allowed to move within the borders of their principality, but they were not allowed to leave it (Fraser, 1992). Nomadic life contributed to the separation from the rest of the population and the preservation of the cultural identity of Roma groups. Throughout history, there have been other attempts to force the Roma to live a sedentary lifestyle, but they have not entirely succeeded, although in the last half century more and more Roma are switching to it. The field research of the Institute for Anthropological Research from 2005/2006, showed that 88.8% of the surveyed Roma were born in the place where they live, as were 69.4% of their parents (Skaric-Juric et al., 2007).

In Croatia, the Roma were first mentioned in Dubrovnik in 1362, and there is a record of their presence in the area of the city of Zagreb already in 1373. Another record that speaks of the presence of Roma in Medjimurje dates from the 17th century. The Roma in Croatia are divided into two basic groups according to their linguistic characteristics and migration patterns – Vlax Roma and Balkan Roma.

Vlax Roma is the name for the groups that share a history of 500 years of slavery in what is now Romania. Long-term slavery had a significant impact on the Vlax Roma and their social structure, as it broke the once homogeneous population into many smaller subgroups that began to differ from each other in terms of customs and way of life. These changes in the social structure were also accompanied by extreme endogamy, which led to changes in their genetic structure (Chaix et al., 2004). In Croatia, the most represented of these Vlax groups are Bayash Roma. The Vlax Roma were not allowed to use their language during slavery, and today their descendants speak the archaic old Romanian language - ljimb d'bjaš. Linguistic research has shown that there are three different dialects of the Bayash Romani language in Croatia. The Erdelj dialect is used by the Roma from Medjimurje, the Baranja muten is characteristic of the Roma from Baranja, and the Ludar muten is used by the Roma from the areas of Kutina, Sisak and Slavonski Brod. There is a similarity between the Baranja and Ludar muten dialects, while the Erdelj dialect differs from them significantly (Radosavljevic, 2010). The first large-scale immigration of Vlax Roma to Croatia took place in the 17th century, when the Austro-Hungarian army occupied Wallachia. The Vlax Roma inhabited the area of the lowlands of large rivers - Sava, Drava and Danube. Other large groups of Vlax

Roma, including Bayash group, came to Croatia after the abolition of slavery (Kalaydjieva et al., 2005).

On the territory of the Republic of Croatia, the second largest group are the Balkan Roma, whose ancestors settled in these areas as early as the 11th century, during the Ottoman Empire. In their everyday speech, these Roma use Romani Chib dialects of the Romani language. Linguistic, but also other socio-cultural characteristics, led to the fragmentation of the Roma population into a number of small groups, which led to the founder effect and the bottleneck that resulted in a distinct genetic substructure of the Roma population (Chaix et al., 2004; Fraser, 1992; Gresham et al., 2001).

The first genetic studies of the Roma population were interested in finding answers to three questions: 1) how similar are Roma and Indians, 2) is there a connection of Roma with European majority populations, and 3) what are the relations between Roma groups from different European countries (Kalaydjieva, Calafell, et al., 2001; Kalaydjieva, Gresham, et al., 2001). Modern genetic research has confirmed the 200-year-old assumptions of linguists (Fraser et al., 1998; Liégeois, 1989) that the Roma originate from an ancestral, proto-Indian population (Kalaydjieva et al., 2005; Malyarchuk et al., 2006; Morar et al., 2004). Research of Y chromosomes (Rai et al., 2012), and of the whole genome (Moorjani et al., 2013), place their origin in the area of northwestern India, which were the original assumptions of the first linguistic and historical studies.

Genetic research has provided insight into the relations of Roma populations with the surrounding majority populations. The Roma showed a certain degree of admixture with the surrounding majority population, but they preserved their genetic and socio-cultural isolation from the majority population to the greatest extent. Martinović Klarić and colleagues found that up to 50.3% of Roma men in Croatia carry Y chromosome haplotypes that point to Indian origin (Klaric et al., 2009), while Peričić Salihović and colleagues found that 26.5% of Roma women and men carry the mitochondrial haplogroup M, which is specific to the Indian subcontinent (Salihovic et al., 2011). Due to genetic drift, limited gene flow from surrounding majority populations and the founder effect during the formation of new Roma groups, there is a relative homogeneity of the genetic makeup within the Roma groups present in Europe today. However, despite this, there is a clear distinction between different Roma groups. (Chaix et al., 2004; Gresham et al., 2001; Kalaydjieva et al., 2005). The results of the mitochondrial DNA research of some of the European Roma groups clearly indicated

different migration routes and patterns of individual Roma groups, specifically the separation of the Vlax Roma from the Balkan and Western European Roma (Salihovic et al., 2011).

All genetic studies conducted so far, starting with the research of Gresham and colleagues in 2001 suggest that modern day Roma descended from a small number of founders who separated from the original population, and later split into numerous other subgroups that gradually began to diverge from one another (Gresham et al., 2001; Kalaydjieva et al., 2005).

In the last few years, the Roma have been in the focus of pharmacogenomic research due to their specific demographic history (Neus Font-Porterias et al., 2021). ADME genes that have been studied so far in Roma groups from different countries are: *ABCB1* gene (Sipeky et al., 2011; Zajc Petranovic et al., 2019), *CYP2B6* gene (Dlouhá et al., 2020; Tomas et al., 2017; Weber et al., 2015), *NAT* genes (Stojanović Marković, Zajc Petranović, Škobalj, et al., 2022; Teixeira et al., 2015), *CYP2C19* gene (Petrović et al., 2019; Sipeky et al., 2013; Teixeira et al., 2015; Zajc Petranovic et al., 2018), *SLCO1B1* gene (Nagy et al., 2015), *VKORC1* gene (Sipeky et al., 2009) and *CYP2D6* gene (Dlouhá et al., 2020; Petrović et al., 2019; Stojanovic Markovic et al., 2022; Weber et al., 2015). In addition to the aforementioned studies of the role of individual genes, Škarić-Jurić and colleagues conducted a study of genome-wide variations in 95 ADME core genes associated with drug response in the Croatian Roma population. Their results confirmed specific positions of isolated populations within the big picture of global ADME distribution. They showed that pharmaco-therapeutic practice in an isolated population cannot be based on pharmacogenetics guidelines for majority populations, but instead has to take their specific genetic profile in consideration (Skaric-Juric et al., 2018).

1.7. RESEARCH PROBLEM AND SCOPE OF THE THESIS

The hypothesis of this study is that, based on the specific socio-cultural characteristics of the Roma, with a high degree of the practice of endogamy being the most prominent one, there are differences in pharmacogenetic profile between different Roma groups, as well as between the Roma and other world populations.

The aim of the research is to determine how socio-cultural and migrational features affected pharmacogenetic phenotype of Roma in Croatia through *CYP2D6* gene analysis. This goal will be met through following specific goals:

1) Analysis of population structure through individual polymorphic loci and haplotypes
2) Haplotype age determination
3) Determination of pharmacogenetic information

The focus of this dissertation is the analysis of the *CYP2D6* gene in the Roma population from Croatia, which belongs to three different migration and dialectal groups. *CYP2D6* is one of the most polymorphic genes responsible for absorption, distribution, metabolism and excretion (ADME), the variations of which affect its enzyme activity. It is located on chromosome 22q13.1 in tandem with the *CYP2D7P* and *CYP2D8P* pseudogenes, at the 3' end of the *CYP2D* cluster. There are over a hundred *CYP2D6* genetic variations and subvariations that differ in single nucleotide polymorphisms (SNPs) or copy number variations (CNV), the latter arising from deletions or multiplications of the *CYP2D6* gene. This gene encodes a homonymous phase I drug-metabolizing enzyme whose metabolic activity is affected by genetic variations in *CYP2D6*. Enzyme activity can vary from complete absence to increased activity and can be grouped into four different phenotypes of drug metabolism and into two subclasses: a) normal metabolizer (NM), b) intermediate metabolizer (IM), c) ultra-rapid metabolizer (UM), and d) poor metabolizer (PM), and subclasses: a) IM through NM, b) PM through IM. The CYP2D6 enzyme is responsible for the metabolism of up to 25% of clinically prescribed drugs, although it accounts for only 2-4% of the total CYP content in the liver.

Regulatory elements of the *CYP2D6* gene region (promoter and enhancer) are believed to play a role in its activity. A promoter is a region of DNA upstream of a gene where proteins such as RNA polymerase and transcription factors bind to initiate transcription. Enhancer

activates promoters at certain locations, times, and levels. Only a few regulatory variants of the *CYP2D6* gene have been studied so far.

Despite their important and proven functional role, knowledge about the distribution of ADME genes in isolated populations is limited. One example of isolated populations is the Roma. They are a transnational minority present in many countries of the world. They originated in India, but migrated to Europe around the 11th century via Central Asia, the Middle East and present-day Turkey. Around 15 million Roma live in the world, of which 12 million live in Europe. Roma are one of the 22 minority populations recognised in Croatia. Roma in Croatia are part of two linguistic categories: Vlax and Balkan Roma. The Vlax Roma use a specific archaic Romani language - ljimb d'bjaš, because they are descendants of Roma who were in slavery for 500 years, and the Balkan Roma use the romani chib language. Gradual migrations over the centuries along with centuries of socio-cultural isolation shaped the Roma population, so in line with the aim of the thesis we wanted to see how their genetic history affected their pharmacogenetics phenotype. This research describes for the first time the SNP variation within the *CYP2D6* gene in a Roma population. To our knowledge, this is the most extensive investigation of this important pharmacogene in this isolated population and it clarifies metabolic phenotypes in three groups of Croatian Roma. Linkage disequilibrium values between regulatory region SNPs and star-defined SNPs were also found and may help predict CYP2D6 activity with greater accuracy. This research not only provides a scientific contribution to the pharmacogenetic knowledge of isolated populations, but can also contribute to clinical application, as it can enable better use of drugs metabolised by CYP2D6 enzyme in the Roma population.

2. LIST OF PUBLICATIONS

2.1. UNTANGLING SNP VARIATIONS WITHIN *CYP2D6* GENE IN CROATIAN ROMA

Journal of *Personalized Medicine*

MDPI

Article

Untangling SNP Variations within *CYP2D6* Gene in Croatian Roma

Anita Stojanović Marković [1], Matea Zajc Petranović [1], Željka Tomas [2], Borna Puljko [3,4], Maja Šetinc [1], Tatjana Škarić-Jurić [1] and Marijana Peričić Salihović [1,*]

[1] Institute for Anthropological Research, 10000 Zagreb, Croatia; astojanovic@inantro.hr (A.S.M.); matea@inantro.hr (M.Z.P.); maja.setinc@inantro.hr (M.Š.); tanja@inantro.hr (T.Š.-J.)
[2] Department for Translational Medicine, Srebrnjak Children's Hospital, 10000 Zagreb, Croatia; ztomas@bolnica-srebrnjak.hr
[3] Croatian Institute for Brain Research, School of Medicine, University of Zagreb, 10000 Zagreb, Croatia; borna.puljko@mef.hr
[4] Department for Chemistry and Biochemistry, School of Medicine, University of Zagreb, 10000 Zagreb, Croatia
* Correspondence: mpericic@inantro.hr

Abstract: *CYP2D6* is a highly polymorphic gene whose variations affect its enzyme activity. To assess whether the specific population history of Roma, characterized by constant migrations and endogamy, influenced the distribution of alleles and thus phenotypes, the *CYP2D6* gene was sequenced using NGS (Next Generation Sequencing) method-targeted sequencing in three groups of Croatian Roma (N = 323) and results were compared to European and Asian populations. Identified single nucleotide polymorphisms (SNPs) were used to reconstruct haplotypes, which were translated into the star-allele nomenclature and later into phenotypes. A total of 43 polymorphic SNPs were identified. The three Roma groups differed significantly in the frequency of alleles of polymorphisms 6769 A > G, 6089 G > A, and 5264 A > G ($p < 0.01$), as well as in the prevalence of the five most represented star alleles: **1*, **2*, **4*, **10*, and **41* ($p < 0.0001$). Croatian Roma differ from the European and Asian populations in the accumulation of globally rare SNPs (6089 G > A, 4589 C > T, 4622 G > C, 7490 T > C). Our results also show that demographic history influences SNP variations in the Roma population. The three socio-culturally different Roma groups studied differ significantly in the distribution of star alleles, which confirms the importance of a separate study of different Roma groups.

Keywords: *CYP2D6*; ADME; pharmacogenetics; population genetics; star allele; Roma; Croatia

Citation: Stojanović Marković, A.; Zajc Petranović, M.; Tomas, Ž.; Puljko, B.; Šetinc, M.; Škarić-Jurić, T.; Peričić Salihović, M. Untangling SNP Variations within *CYP2D6* Gene in Croatian Roma. *J. Pers. Med.* **2022**, *12*, 374. https://doi.org/10.3390/jpm12030374

Academic Editor: Chiara Villa

Received: 19 January 2022
Accepted: 23 February 2022
Published: 28 February 2022

1. Introduction

The *CYP2D6* gene encodes the phase I drug-metabolizing homonymous enzyme and is located in tandem with pseudogenes *CYP2D7P* and *CYP2D8P* on chromosome 22q13.1, at the 3′ end of the *CYP2D* cluster [1,2]. It contains nine exons consisting of 1461 codons and is highly polymorphic, with more than a hundred genetic variations and numerous subvariants that differ in single nucleotide polymorphisms (SNPs) or copy number variations (CNV); the latter resulting from *CYP2D6* gene deletion or multiplication. In 1996, a group of international experts in pharmacogenetics decided to systematize allelic variants of *CYP2D6* proposing a haplotype-based star (*) nomenclature system [3]. Since then, mostly due to the development of DNA sequencing technology, a tremendous amount of allelic and suballelic variants have been identified and classified by the Pharmacogene Variation Consortium [4].

The most important consequence of *CYP2D6* genetic variations is their influence on the metabolizing activity of the CYP2D6 enzyme. These variations were broadly grouped into four different drug-metabolic phenotypes of the CYP2D6 enzyme; (1) poor metabolizer (PM—only null activity alleles detected), (2) intermediate metabolizer (IM—one normal activity allele with

one null activity allele; or two decreased activity alleles), (3) extensive metabolizer (EM—two normal activity alleles; or a combination of one increased activity allele with one decreased activity allele), and (4) ultra-rapid metabolizer (UM—a combination of one normal activity allele with one increased activity allele) [5–8]. Metabolizing activity depends not only on the genotype, but is also influenced by a number of physiological, pathological, and environmental factors [9]. CYP2D6 is involved in the metabolism of up to 25% of drugs commonly used in medicine [10], although it constitutes only 2–4% of the total CYP content in the liver [11,12]. CYP2D6 metabolizes a wide range of drugs: antiarrhythmics, tricyclic and second-generation antidepressants, β-blockers, anti-cancer drugs, several opioid analgesics including codeine and tramadol, and many more [13–18]. Additionally, variations in the *CYP2D6* gene have been studied as a risk factor for a number of diseases: Parkinson's disease [19–21], schizophrenia and other psychiatric diseases [15,22], Alzheimer's disease [23,24], and several forms of cancer [25,26].

The role of CYP2D6 in the metabolism of natural xenobiotics has been studied scarcely, but this enzyme is known to have a very high affinity for alkaloids [27]. Therefore, its evolutionary role is thought to be related to alkaloid metabolism in food. There is a hypothesis that due to food constraints relative to the size of the population in Northwest Africa some 10,000–20,000 years ago, a selection pressure occurred that favored the survival of subjects capable of detoxifying plant toxins to a higher extent, increasing the number of plants that can provide useful food [28]. Changes in diet throughout human history have exerted selective pressure on genes whose products metabolize food compounds, the best example being N-acetyltransferase 2 (e.g., [29]). Fuselli (2010) suggested that current patterns of genetic diversity in *CYP2D6* are the result of selective pressure imposed by new or more concentrated CYP2D6 substrates that emerged in food production especially at the beginning of the Neolithic transition, in the presence of poorer nutritional conditions and higher disease burdens [30].

Genes encoding phase I and phase II metabolizing enzymes, drug transporters, and modifiers show a population-specific diversity, indicating that their variability has been shaped by evolutionary mechanisms [31]. The global distribution of *CYP2D6* allele frequencies is differing. Some alleles are present at similar frequencies across the world, while others are ethnically or geographically specific. The nonfunctional allele *CYP2D6*4* is the most frequently found in Europe, the decreased-function *CYP2D6*10* allele is mostly present in Asia and East Asia, alleles *CYP2D6*17* and **29* are characteristic for Africa and Afro-descendant populations, while *CYP2D6*41* allele and gene amplifications are the most common in Middle Eastern populations [4,32,33]. Despite the proven functional role of CYP2D6 enzyme and its extensive research worldwide, knowledge of *CYP2D6* allele frequencies within isolated populations like the Roma population, Basques, Ashkenazi Jews, and Saami is very limited [34].

The Roma (Gypsy) population is a transnational minority present in many countries of the world. They originated in India and arrived in Europe around the 11th century via central Asia (Afghanistan and Persia), the Middle East, and present-day Turkey. The Roma population is estimated at 15 million people, of whom 12 million live in Europe. Roma in Croatia belong to two socio-culturally and linguistically different groups: Vlax Roma, who speak ljimb'd bayash language, and Balkan Roma, who speak dialects of romani chib language. The Vlax Roma are descendants of the Roma who crossed the Danube River between the 13th and 15th century and upon arrival in Wallachia, Transylvania (both in present-day Romania) and Moldavia were enslaved to work in the mines for the next 500 years. During that time, they were forbidden to use their own language, so their descendants are recognized by a specific archaic Romanian language. In the 19th century, after slavery was abolished, the Vlax Roma started a new migration wave; after leaving Romania, they settled in Hungary, Croatia, Serbia, and other Balkan states, as well as in other parts of Europe, while some even reached the United States [35,36]. Balkan Roma in Croatia are descendants of the Roma who arrived in the Balkans in the 11th century, and during the Ottoman expansion, some groups of these Balkan Roma moved further west. Socio-cultural characteristics of the Roma population, such as

strict rules of endogamy, have led to the founder and the bottleneck effects that have caused a genetic structure of Roma to differ compared to other populations [35,37,38], which has been shown to affect ADME (Absorption, Distribution, Metabolism and Excretion) genes variations as well [39]. In Croatia, the Vlax Roma mostly inhabit the Baranja Region, in the eastern part of Croatia, and the Medjimurje Region, in the north of Croatia. The Balkan Roma mostly live in Zagreb area.

The main objective of this study was to estimate variation within the *CYP2D6* gene among three socio-culturally and geographically distinct Croatian Roma groups (Balkan Roma, and Vlax Roma from the Baranja region and Medjimurje region), to find out whether their specific population history influenced the distribution of *CYP2D6* alleles and consequently phenotypes.

2. Materials and Methods

We analyzed 323 DNA samples, all collected during field studies of the ongoing multidisciplinary anthropological, molecular-genetic, and epidemiological investigations of Roma groups in Croatia. Samples were collected during field research in Vlax Roma settlements in the Baranja region and Medjimurje region and in Balkan Roma settlements in the city of Zagreb, Croatia (Figure 1). All respondents participated in the research voluntarily, and with the help of Roma volunteers, were introduced to the objectives, methods, and the anticipated contribution of the project. The protocol of the study was approved by the Scientific Board and the Ethics Committee of the Institute for Anthropological Research in Zagreb, Croatia.

Figure 1. Sampling locations in Croatia.

DNA was extracted from peripheral blood using the salting-out method [40]. The entire *CYP2D6* gene (ENSG00000100197 chromosome 22: 42,126,499–42,130,881, GRCh38.p12) was sequenced with Illumina MiSeq V3 kit using the method Genotyping-in-Thousands by sequencing (GT-seq) in a commercial laboratory. GT-seq is a multiplexed amplicon sequencing method that allows for the simultaneous genotyping of hundreds of SNPs across thousands of individuals in a single library, making library preparation simple and cost-effective [41]. Raw reads were demultiplexed and sequencing adapter remnants were clipped from all reads (reads with final length < 100 bases were discarded). Primer sequences were removed, and sequence fragments were turned into forward-reverse primer orientation. Low-quality reads were discarded (Phred quality score of 20 over a window of 10 bases). Quality trimmed reads were aligned against all clusters using BWA version 0.7.12.

The FreeBayes v1.02-16 was used for variant discovery and genotyping of the samples. Variants were filtered using a set of GT-seq specific rules (minimum allele count must exceed eight reads, minimum allele frequency across all samples must exceed 10%).

Allele and genotype frequencies and Hardy–Weinberg equilibrium were calculated using VCFtools [42]. Differences in genotype and allele frequencies between Roma groups were tested using the Chi-square test and Fisher's exact test. The analyses were performed using the SPSS Statistics 21.0 statistical package for Windows (SPSS Inc., Chicago, IL, USA) and *p*-values were corrected for multiple testing using Bonferroni correction.

CYP2D6 gene sequencing identified 43 single nucleotide polymorphisms (SNPs), 28 of which were used for haplotype reconstruction. The 14 SNPs were excluded due to low level of heterozygosity (rs28371732, rs867985262, rs79596243, rs28578778, rs141009491, rs762158210, rs35742686, rs28371717, rs1349481801, rs376056664, rs189736703, rs28371703, rs1080992, and rs769811346). The low level of heterozygosity was considered for cases in which there was no minor allele in two Croatian Roma groups, while in the third Roma group it occurred with a frequency of two or less. The remaining insertion polymorphism (rs1269631565) was excluded because it cannot be phased by Phase software due to its indel nature. However, this polymorphism is not among star alleles' defining SNPs.

Haplotypes of the Croatian Roma groups were inferred using Phase ver. 2.1 [43,44]. Haplotypes were translated into the star allele nomenclature using the data provided at the PharmVar website [45]. Star diplotypes were translated to phenotypes and classified into three metabolizer categories—normal, intermediate, and poor according to the guidelines [46].

Software Arlequin 3.5 [47] was used to infer intra-population diversity indexes (haplotype and nucleotide diversity) and population pairwise F_{ST} values, statistical significance assessed by generating 100,000 random samples. Possible departure from selective neutrality was tested using the Ewens–Watterson (EW) homozygosity test also implemented in Arlequin. Statistical significance was assessed by generating 10,000 random samples under the null conditions of no selection and constant population size.

The relations among haplotypes were shown in networks, which were calculated by the median-joining (MJ) procedure with default settings [48] using the program NETWORK 10.2.0.0 [49].

The inter-population comparisons of the Roma with other populations, based on frequencies of minor alleles, were performed using the data from the gnomAD database (v3.1.2) for East Asian, South Asian, and European (non-Finnish) populations. A comparison of the star allele distribution between the surveyed Roma groups and world population divided into major ethnic groups was performed using data from studies listed in Supplementary Table S2 in the work of Gaedigk (2017) [4]. Data were visualized in six graphs, each one representing a star allele found in more than 5% of the Roma in the Croatian Roma groups. Because the major ethnic groups consisted of data from multiple studies, the average star allele frequency for each major ethnicity group was calculated by weighting according to the sample size of each study.

3. Results

Sequencing of the *CYP2D6* gene in DNA samples obtained from three Croatian Roma groups (323 persons in total) reveal 43 polymorphic positions, which are listed in Table 1. Due to low heterozygosity (absence of heterozygotes and recessive homozygotes), only 21 SNPs could be tested for HWE in all three Croatian Roma groups. The remaining 22 SNPs were tested for HWE in only one or two Croatian Roma groups. Most of the SNPs which were polymorphic in all three Roma groups were in Hardy–Weinberg equilibrium (HWE); after applying Bonferroni correction, exceptions were intron variants 8602 A > G (rs2004511) in total Roma sample, 6188 G > A (rs1081004) in all three Roma groups and in the total Roma sample and noncoding transcript exon 5264 A > G (rs29001678) in total Roma sample.

Table 1. Genotype and allele frequencies of *CYP2D6* polymorphisms in the three Croatian Roma samples (Balkan, Medjimurje, Baranja) and in the combined sample.

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Cro-Roma	X^2	*p*
9200 G > C (rs1135840) missense variant	ultrarapid metabolism of debrisoquine, deutetrabenazine response, tamoxifen response, tramadol response, benign	genotype	G/G	8 (8.16%)	20 (18.52%)	14 (11.97%)	42 (13.00%)	0.8873	0.1119	0.6276	0.3772	18.22617	0.0991
			C/C	51 (52.04%)	44 (40.74%)	46 (39.32%)	141 (43.65%)						
			G/C	39 (39.80%)	43 (39.81%)	56 (47.86%)	138 (42.72%)						
		allele	C	141 (71.94%)	131 (61.21%)	148 (63.79%)	420 (65.42%)					5.62592	0.06003
8848 G > A (rs28371732) synonymous variant		genotype	G/G	96 (97.96%)	102 (94.44%)	92 (78.63%)	290 (89.78%)		0.9605		0.9766	1.83154	0.40021
			G/A	0	1 (0.93%)	0	1 (0.31%)						
		allele	A	0	1 (0.48%)	0	1 (0.17%)					1.82838	0.40084
8810 C > T (rs4987144) intron variant		genotype	C/C	44 (44.90%)	59 (54.63%)	59 (50.43%)	162 (50.15%)	0.8366	0.0538	0.066	0.0363	3.49495	0.47865
			T/T	10 (10.20%)	13 (12.04%)	16 (13.68%)	39 (12.07%)						
			C/T	44 (44.90%)	36 (33.33%)	42 (35.90%)	122 (37.77%)						
		allele	T	64 (32.65%)	62 (28.70%)	74 (31.62%)	200 (30.96%)					0.82556	0.66181
8604 G > A (rs28371730) intron variant		genotype	G/G	40 (40.82%)	45 (41.67%)	50 (42.74%)	135 (41.80%)	0.7361	0.8758	0.2036	0.3834	1.12704	0.88996
			A/A	14 (14.29%)	13 (12.04%)	19 (16.24%)	46 (14.24%)						
			G/A	44 (44.90%)	50 (46.30%)	48 (41.03%)	142 (43.96%)						
		allele	A	72 (36.73%)	76 (35.19%)	86 (36.75%)	234 (36.22%)					0.15128	0.92175
8602 A > G (rs2004511) intron variant		genotype	A/A	37 (37.76%)	30 (27.78%)	45 (38.46%)	112 (34.67%)	0.3336	0.0007	0.0088	**0.0001**	5.54417	0.23588
			G/G	11 (11.22%)	9 (8.33%)	7 (5.98%)	27 (8.36%)						
			A/G	50 (51.02%)	69 (63.89%)	65 (55.56%)	184 (56.97%)						
		allele	G	72 (36.73%)	87 (40.28%)	79 (33.76%)	238 (36.84%)					2.05158	0.35851
8565 dup (rs1269631565) intron variant		genotype	T/T	94 (95.92%)	86 (79.63%)	106 (90.60%)	286 (88.54%)	0.8366	0.2386	0.5936	0.2749	14.2025	**0.00082**
			TT/TT	0	0	0	0						
			T/TT	4 (4.08%)	22 (20.37%)	11 (9.40%)	37 (11.46%)						
		allele	TT	4 (2.04%)	22 (10.19%)	11 (4.70%)	37 (5.73%)					13.3396	0.00127
8504 G > A (rs867985262) intron variant		genotype	G/G	98 (100.00%)	107 (99.07%)	117 (100.00%)	322 (99.69%)		0.9614		0.9778	1.99692	0.36845
			G/A	0	1 (0.93%)	0	1 (0.31%)						
		allele	A	0	1 (0.46%)	0	1 (0.15%)					1.99383	0.36902

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Croatia-Roma	X^2	*p*
8498 A > G (rs79596243) intron variant		genotype	A/A	98 (100.00%)	107 (99.07%)	117 (100.00%)	322 (99.69%)		0.9614		0.9778	1.99692	0.36845
			A/G	0	1 (0.93%)	0	1 (0.31%)						
		allele	G	0	1 (0.46%)	0	1 (0.15%)					1.99383	0.36902
8455 C > A (rs28371729) intron variant	tramadol response	genotype	C/C	96 (97.96%)	107 (99.07%)	116 (99.15%)	319 (98.76%)	0.9187	0.9614	0.9630	0.9108	0.74298	0.68971
			C/A	2 (2.04%)	1 (0.93%)	1 (0.85%)	4 (1.24%)						
		allele	A	2 (1.02%)	1 (0.46%)	1 (0.43%)	4 (0.62%)					0.73835	0.69131
8413 T > C (rs28578778) intron variant		genotype	T/T	98 (100.00%)	108 (100.00%)	116 (99.15%)	322 (99.69%)			0.9630	0.9778	1.76615	0.41351
			T/C	0	0	1 (0.85%)	1 (0.31%)						
		allele	C	0	0	1 (0.43%)	1(0.15%)					1.76341	0.41408
8404 A > C (rs1985842) intron variant		genotype	A/A	8 (8.16%)	19 (17.59%)	13 (11.11%)	40 (12.38%)	1.000	0.2143	0.4040	0.7101	7.04036	0.13377
			C/C	50 (51.02%)	44 (40.74%)	46 (39.32%)	140 (43.34%)						
			A/C	40 (40.82%)	45 (41.67%)	58 (49.57%)	143 (44.27%)						
		allele	C	140 (71.43%)	133 (61.57%)	150 (64.10%)	423 (65.48%)					4.72262	0.0943
8199 C > T (rs200335621) synonymous variant		genotype	C/C	96 (97.96%)	106 (98.15%)	107 (91.45%)	309 (95.67%)	0.9187	0.9226	0.6292	0.6905	7.85586	0.01968
			C/T	2 (2.04%)	2 (1.85%)	10 (8.55%)	14 (4.33%)						
		allele	T	2 (1.02%)	2 (0.93%)	10 (4.27%)	14 (2.17%)					7.68184	0.02147
8180 G > C (rs141009491) missense variant		genotype	G/G	98 (100.00%)	108 (100.00%)	116 (99.15%)	322 (99.69%)			0.9630	0.9778	1.76615	0.41351
			G/C	0	0	1 (0.85%)	1 (0.31%)						
		allele	C	0	0	1 (0.43%)	1 (0.15%)					1.76341	0.41408
8008 G > A (rs28371725) intron variant	deutetrabenazine response, tamoxifen response, tramadol response	genotype	G/G	66 (67.35%)	88 (81.48%)	75 (64.10%)	229 (70.90%)	0.0543	0.4254	0.9517	0.1832	13.69378	0.00834
			A/A	5 (5.10%)	0 (0.00%)	4 (3.42%)	9 (2.79%)						
			G/A	20 (20.41%)	15 (13.89%)	34 (29.06%)	69 (21.36%)						
		allele	A	30 (16.48%)	15 (7.28%)	42 (18.58%)	87 (14.17%)					12.0599	0.00241

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Croatia-Roma	X^2	*p*
7870 C > T (rs16947) missense variant	benign, ultrarapid metabolism of debrisoquine, deutetrabenazine response, tamoxifen response, tramadol response	genotype	C/C	34 (34.69%)	46 (42.59%)	48 (41.03%)	128 (39.63%)	0.0511	0.6876	0.0393	0.0236	7.05147	0.13319
			T/T	21 (21.43%)	10 (9.26%)	21 (17.95%)	52 (16.10%)						
			C/T	35 (35.71%)	47 (43.52%)	42 (35.90%)	124 (38.39%)						
		allele	T	77 (42.78%)	67 (32.52%)	84 (37.84%)	228 (37.5%)					4.32612	0.11497
7632_7634 del (rs762158210) inframe deletion		genotype	GAGAA/GAGAA	67 (68.37%)	37 (34.26%)	23 (19.66%)	127 (39.32%)	0.9028			0.9293	1.76652	0.41343
			GAGAA/GA	2 (2.04%)	0	0	2 (0.62%)						
		allele	GA	2 (1.45%)	0	0	2 (0.78%)					1.75272	0.4163
7569 del (rs35742686) frameshift variant	poor metabolizer of debrisoquine	genotype	CAG/CAG	92 (93.88%)	97 (89.81%)	95 (81.20%)	284 (87.93%)	0.9584			0.9763	2.07179	0.35491
			CAG/CG	1 (1.02%)	0	0	1 (0.31%)						
		allele	CG	1 (0.54%)	0	0	1 (0.18%)					2.06814	0.3556
7503 G > T (rs28371717) missense variant	tramadol response	genotype	G/G	98 (100.00%)	108 (100.00%)	116 (99.15%)	322 (99.69%)			0.9630	0.9778	1.76615	0.41351
			G/T	0	0	1 (0.85%)	1 (0.31%)						
		allele	T	0	0	1 (0.43%)	1 (0.15%)					1.76341	0.41408
7490 T > C (rs17002852) synonymous variant	tramadol response	genotype	T/T	91 (92.86%)	108 (100.00%)	104 (88.89%)	303 (93.81%)	0.7139		0.3395	0.2439	12.93642	0.01159
			C/C	0	0	1 (0.85%)	1 (0.31%)						
			T/C	7 (7.14%)	0	12 (10.26%)	19 (5.88%)						
		allele	C	7 (3.57%)	0	14 (5.98%)	21 (3.25%)					12.87541	0.0016
7117 A > G (rs2267447) intron variant	tramadol response	genotype	A/A	53 (54.08%)	63 (58.33%)	63 (53.85%)	179 (55.42%)	0.4295	0.3927	0.2320	0.4934	3.98125	0.40855
			G/G	9 (9.18%)	4 (3.70%)	5 (4.27%)	18 (5.57%)						
			A/G	36 (36.73%)	41 (37.96%)	49 (41.88%)	126 (39.01%)						
		allele	G	54 (27.55%)	49 (22.68%)	59 (25.21%)	162 (25.08%)					2.08793	0.35206

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Cro-Roma	X^2	*p*
6866 G > A (rs3892097) splice acceptor variant	amitriptyline response, antidepressants response—dosage. toxicity/ADR, clomipramine response, poor metabolizer of debrisoquine, deutetrabenazone response, tamoxifen response, tramadol response, desipramine response, doxepine response, imipramine response, nortriptyline response, trimipramine response, urinary metabolite levels in chronic kidney disease	genotype	G/G	57 (58.16%)	74 (68.52%)	71 (60.68%)	202 (62.54%)	0.5398	0.0522	0.0974	0.1578	11.76100	0.01922
			A/A	7 (7.14%)	0	2 (1.71%)	9 (2.79%)						
			G/A	34 (34.69%)	34 (31.48%)	44 (37.61%)	112 (34.67%)						
		allele	A	48 (24.49%)	34 (15.74%)	48 (20.51%)	130 (20.12%)					4.92789	0.0851
6769 A > G (rs1135824) missense variant	likely benign, germline origin	genotype	A/A	92 (93.88%)	108 (100.00%)	117 (100.00%)	317 (98.14%)	0.7546			0.8662	14.03625	**0.00090**
			A/G	6 (6.12%)	0	0	6 (1.86%)						
		allele	G	6 (3.06%)	0	0	6 (0.93%)					13.90466	**0.00096**
6684 C > T (rs1349481801) synonymous variant		genotype	C/C	98 (100.00%)	108 (100.00%)	116 (99.15%)	322 (99.69%)			0.9630	0.9778	1.76615	0.41351
			C/T	0	0	1 (0.85%)	1 (0.31%)						
		allele	T	0	0	1 (0.43%)	1 (0.15%)					1.76341	0.41408
6681 G > C (rs1058164) synonymous variant		genotype	G/G	8 (8.16%)	19 (17.59%)	13 (11.11%)	40 (12.38%)	1.000	0.2803	0.4040	0.7749	7.07838	0.13180
			C/C	50 (51.02%)	43 (39.81%)	46 (39.32%)	139 (43.03%)						
			G/C	40 (40.82%)	46 (42.59%)	58 (49.57%)	144 (44.58%)						
		allele	C	140 (71.43%)	132 (61.11%)	150 (64.10%)	422 (65.33%)					5.07114	0.0792
6460 T > C (rs376056664) intron variant		genotype	T/T	96 (97.96%)	105 (97.22%)	115 (98.29%)	316 (97.83%)	0.9593			0.9776	2.2752	0.3206
			C/C	0	0	0	0						
			T/C	1 (1.02%)	0	0	1 (0.31%)						
		allele	C	1 (0.52%)	0	0	1 (0.16%)					2.27162	0.32116

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Croroma	X^2	*p*
6313 G > A (rs189736703) intron variant		genotype	G/G	98 (100.00%)	108 (100.00%)	114 (97.44%)	320 (99.07%)			0.9626	0.9777	1.79690	0.40720
			G/A	0	0	1 (0.85%)	1 (0.31%)						
		allele	A	0	0	1 (0.43%)	1 (0.16%)					1.7941	0.40777
6188 G > A (rs1081004) intron variant	tramadol response	genotype	G/G	94 (95.92%)	103 (95.37%)	110 (94.02%)	307 (95.05%)	0.0001	**0.00002**	**$<10^{-5}$**	**$<10^{-5}$**	2.57173	0.63184
			A/A	1 (1.02%)	3 (2.78%)	5 (4.27%)	9 (2.79%)						
			G/A	3 (3.06%)	2 (1.85%)	2 (1.71%)	7 (2.17%)						
		allele	A	5 (2.56%)	8 (3.70%)	12 (5.13%)	25 (3.87%)					1.92838	0.38129
6089 G > A (rs368389952) intron variant		genotype	G/G	86 (87.76%)	108 (100.00%)	117 (100.00%)	311 (96.28%)	**$<10^{-5}$**			**$<10^{-5}$**	28.61408	**$<10^{-5}$**
			A/A	6 (6.12%)	0	0	6 (1.86%)						
			G/A	6 (6.12%)	0	0	6 (1.86%)						
		allele	A	18 (9.18%)	0	0	18 (2.79%)					42.51105	**$<10^{-5}$**
6057 C > T (rs1081003) synonymous variant		genotype	C/C	94 (95.92%)	96 (88.89%)	114 (97.44%)	304 (94.12%)	0.8366	0.5410	0.8883	0.5860	8.23424	0.01629
			C/T	4 (4.08%)	12 (11.11%)	3 (2.56%)	19 (5.88%)						
		allele	T	4 (2.04%)	12 (5.56%)	3 (1.29%)	19 (2.94%)					7.98471	0.01846
6015 C > G (rs28371705) synonymous variant		genotype	C/C	91 (92.86%)	108 (100.00%)	115 (98.29%)	314 (97.21%)	0.7139		0.9257	0.7996	10.4630	0.0054
			G/G	0	0	0	0						
			C/G	7 (7.14%)	0	2 (1.71%)	9 (2.79%)						
		allele	G	7 (3.57%)	0	2 (0.86%)	9 (1.39%)					10.31512	0.00576
6002 A > G (rs28371704) missense variant	tramadol response	genotype	A/A	91 (92.86%)	108 (100.00%)	115 (98.29%)	314 (97.21%)	0.7553		0.9257	0.8214	8.5260	0.0141
			G/G	0	0	0	0						
			A/G	6 (6.12%)	0	2 (1.71%)	8 (2.48%)						
		allele	G	6 (3.09%)	0	2 (0.86%)	8 (1.24%)					8.41879	0.01486

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Croroma	X^2	*p*
5992 C > A (rs28371703) intron variant		genotype	G/G	95 (96.94%)	108 (100.00%)	116 (99.15%)	319 (98.76%)			0.9630	0.9777	1.74048	0.41885
			G/A	0	0	1 (0.85%)	1 (0.31%)						
		allele	A	0	0	1 (0.43%)	1 (0.16%)					1.73776	0.41942
5289 C > T (rs29001678) noncoding transcript exon variant		genotype	C/C	80 (81.63%)	96 (88.89%)	97 (82.91%)	273 (84.52%)	0.9110	0.005	0.001	1.8×10^{-5}	3.88908	0.42123
			T/T	0	1 (0.93%)	2 (1.71%)	3 (0.93%)						
			C/T	2 (2.04%)	0	2 (1.71%)	4 (1.24%)						
		allele	T	2 (1.22%)	2 (1.03%)	6 (2.97%)	10 (1.79%)					2.54616	0.27997
5264 A > G (rs1081000) noncoding transcript exon variant		genotype	A/A	86 (87.76%)	108 (100.00%)	115 (98.29%)	309 (95.67%)	0.5538		0.9257	0.7116	19.53312	**$<10^{-5}$**
			A/G	11 (11.22%)	0	2 (1.71%)	13 (4.02%)						
		allele	G	11 (5.67%)	0	2 (0.85%)	13 (2.02%)					42.511	**$<10^{-5}$**
5119 C > T (rs1065852) missense variant	poor metabolizer of debrisoquine, deutetrabenazone response, tamoxifen response, tramadol response, response to serotonin reuptake inhibitors in major depressive disorder	genotype	C/C	53 (54.08%)	63 (58.33%)	63 (53.85%)	179 (55.42%)	0.2532	0.6703	0.6483	0.8448	3.08176	0.54424
			T/T	10 (10.20%)	5 (4.63%)	7 (5.98%)	22 (6.81%)						
			C/T	35 (35.71%)	40 (37.04%)	47 (40.17%)	122 (37.77%)						
		allele	T	55 (28.06%)	50 (23.15%)	61 (26.07%)	166 (25.70%)					1.32564	0.5154
5101 C > T (rs138100349) missense variant		genotype	C/C	97 (98.98%)	106 (98.15%)	114 (97.44%)	317 (98.14%)	0.9595	0.9226	0.8883	0.8662	0.69712	0.70570
			C/T	1 (1.02%)	2 (1.85%)	3 (2.56%)	6 (1.86%)						
		allele	T	1 (0.51%)	2 (0.93%)	3 (1.28%)	6 (0.93%)					0.69059	0.70801
5050 G > A (rs769258) missense variant	tramadol response, likely benign	genotype	G/G	95 (96.94%)	108 (100.00%)	115 (98.29%)	318 (98.45%)	0.8777		0.9257	0.8885	3.19059	0.20285
			G/A	3 (3.06%)	0	2 (1.71%)	5 (1.55%)						
		allele	A	3 (1.53%)	0	2 (0.85%)	5 (0.77%)					3.1657	0.20539
4818 G > A (rs372204775) intron variant		genotype	G/G	95 (96.94%)	99 (91.67%)	111 (94.87%)	305 (94.43%)	0.9183	0.6514	0.7759	0.6266	4.02766	0.13348
			G/A	2 (2.04%)	9 (8.33%)	6 (5.13%)	17 (5.26%)						
		allele	A	2 (1.03%)	9 (4.17%)	6 (2.56%)	17 (2.64%)					3.91845	0.14097

Table 1. *Cont.*

rsID	Clinical Implications	Genotypes and Alleles		Balkan *N* (%)	Medjimurje *N* (%)	Baranja *N* (%)	Combined *N* (%)	HWE Balkan	HWE Medjimurje	HWE Baranja	HWE Croatia-Roma	X^2	*p*
4666 A > G (rs530422334)intron variant	tramadol response	genotype	A/A	98 (100.00%)	101 (93.52%)	113 (96.58%)	312 (96.59%)		0.7278	0.8508	0.7556	6.56138	0.03760
			A/G	0	7 (6.48%)	4 (3.42%)	11 (3.41%)						
		allele	G	0	7 (3.24%)	4 (1.71%)	11 (1.70%)					6.44772	0.0398
4655 G > A (rs1080992) intron variant		genotype	G/G	98 (100.00%)	106 (98.15%)	117 (100.00%)	321 (99.38%)		0.9226		0.9555	4.00629	0.13491
			G/A	0	2 (1.85%)	0	2 (0.62%)						
		allele	A	0	2 (0.93%)	0	2 (0.31%)					3.99385	0.13575
4623 G > T (rs769811346) intron variant		genotype	G/G	98 (100.00%)	108 (100.00%)	116 (99.15%)	322 (99.69%)			0.9630	0.9778	1.76615	0.41351
			G/T	0	0	1 (0.85%)	1 (0.31%)						
		allele	T	0	0	1 (0.43%)	1 (0.15%)					1.76341	0.41408
4622 G > C (rs374672076) intron variant		genotype	G/G	81 (82.65%)	98 (90.74%)	100 (85.47%)	279 (86.38%)	0.3471	0.6139	0.3968	0.1890	2.98457	0.22486
			G/C	17 (17.35%)	10 (9.26%)	17 (14.53%)	44 (13.62%)						
		allele	C	17 (8.67%)	10 (4.63%)	17 (7.26%)	44 (6.81%)					2.42978	0.29719
4589 C > T (rs566383351) intron variant		genotype	C/C	67 (68.37%)	85 (78.70%)	85 (72.65%)	237 (73.37%)	0.0629	0.2155	0.0866	0.0058	2.85917	0.23941
			C/T	31 (31.63%)	23 (21.30%)	32 (27.35%)	86 (26.63%)						
		allele	T	31 (15.82%)	23 (10.65%)	32 (13.68%)	86 (13.31%)					2.42008	0.29819

HWE (Bonferroni correction) $p = 0.00037$. X^2 (Bonferroni correction) $p = 0.0012$. Significant values are shown in bold.

A total of 21 SNPs were polymorphic in all three Roma groups (Table 1). Chi-square test results showed that four SNPs differed significantly in frequencies of genotypes between the three analyzed groups after Bonferroni's correction: 8565 dup (rs1269631565), 6769 A > G (rs1135824), 6089 G > A (rs368389952), and 5264 A > G (rs1081000). A significant difference in allele frequencies between the three Croatian Roma groups was found in the same SNPs (after Bonferroni's correction): 8565 dup (rs1269631565), 6769 A > G (rs1135824), 6089 G > A (rs368389952), and 5264 A > G (rs1081000).

Furthermore, we compared minor allele frequencies of 43 polymorphic SNPs from this study with their frequencies in East Asian, South Asian, and European populations. The lowest number of polymorphic SNPs was found in East Asian populations, where 18 out of 43 investigated SNPs were monomorphic and 13 SNPs had MAFs less than 1%. In South Asian populations five SNPs were monomorphic and six SNPs had MAFs less than 1%, while only two SNPs were monomorphic in Europeans, but 16 SNPs had MAFs less than 1% (Supplementary Table S1). SNPs found to be polymorphic in the investigated Roma groups had higher MAFs in South Asian than in the European populations. The distribution of allele frequencies of SNPs with MAFs > 0.05 in at least one of the populations is shown in Figure 2.

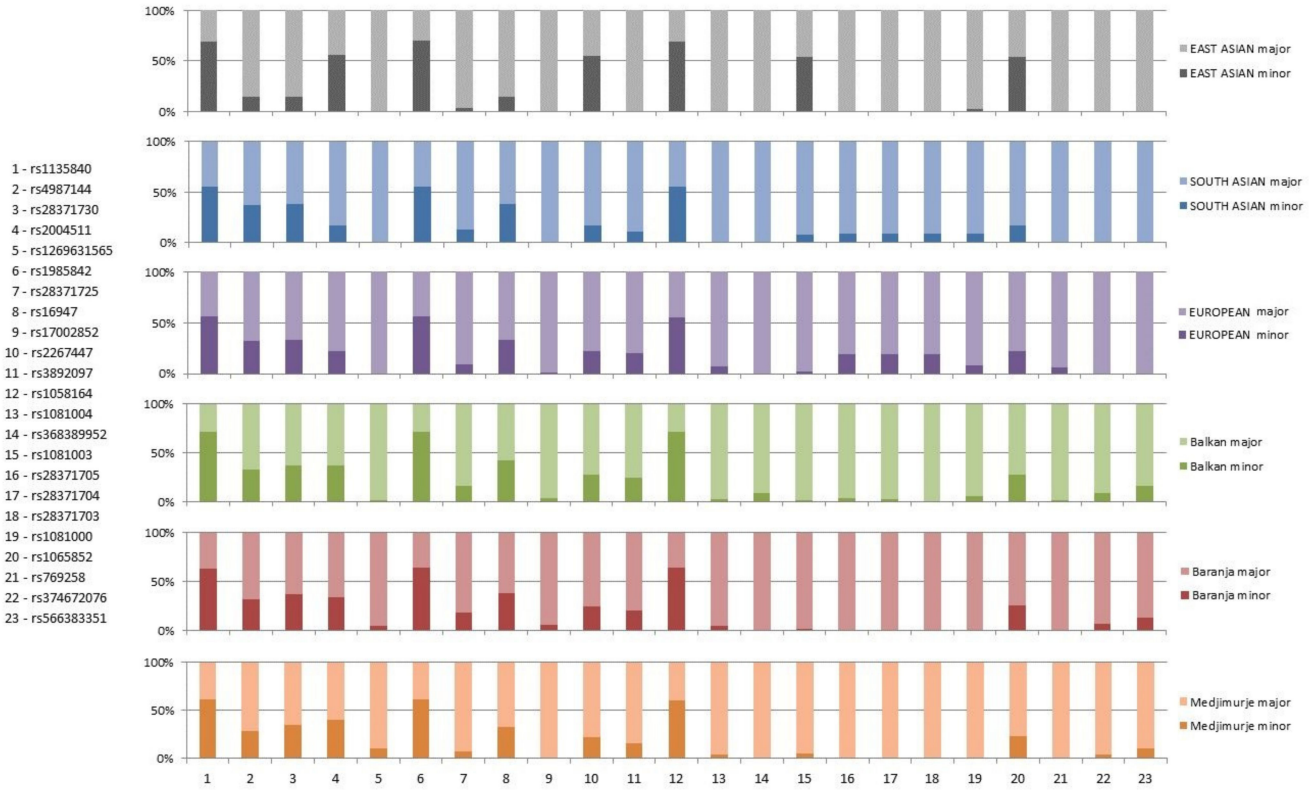


Figure 2. Allele frequencies of *CYP2D6* SNPs in East Asian, South Asian, and European (non-Finnish) populations (as reported in gnomAD database, version 3.1.1) and in three Croatian Roma populations.

A total of 93 *CYP2D6* haplotypes were reconstructed from polymorphic SNPs, as explained in the Materials and Methods section (Supplementary Table S2). Intra-population analysis based on all reconstructed haplotypes revealed the highest diversity in the population of Roma from Baranja and the lowest among Roma from Medjimurje (Table 2). Pairwise population F_{ST} distances showed the highest difference between Balkan and Medjimurje Roma (F_{ST} = 0.01249), followed by similar distances between Roma from Balkan and Baranja (F_{ST} = 0.0028) and between Baranja and Medjimurje (F_{ST} = 0.0027). The exact test of sample differentiation based on haplotype frequencies showed significant differences between samples. In order to exclude the possible influence of selection on *CYP2D6* haplotype distribution, the Ewens–Watterson test of selective neutrality was performed.

All Roma groups have higher observed than expected homozygosity, but results were insignificant which rules out the influence of directional selection.

Table 2. Intra-population diversity and results of the Ewens–Watterson test of selective neutrality.

Roma Group	No. of Polym. Loci	No. of Haplotypes	Haplotype Diversity	Nucleotide Diversity	Observed F Value *	Expected F Value *	*p*-Value *
Balkan	27	46	0.9490	0.2046	0.0558	0.0546	0.6408
Baranja	26	47	0.9154	0.2035	0.0885	0.0574	0.9665
Medjimurje	21	37	0.9114	0.1791	0.0929	0.0762	0.8412

* Ewens–Watterson test of selective neutrality.

A total of 89 haplotypes were translated into star nomenclature, which resulted in 10 star alleles (*) (Table 3). The reference *CYP2D6*1* allele was the most common in each of the three Roma groups individually, and in the entire Roma sample (33.1%). In addition to **1*, the other most prevalent alleles (total sample prevalence larger than 5%) were star alleles **2*, **4*, **10*, and **41*—the five listed star alleles were found in 96.4% of the total Roma sample. Of the five remaining star alleles, two were not found in all three Roma groups, and the *CYP2D6*65* allele was found in only one person from Baranja.

Table 3. Distribution of star alleles in the total Croatian Roma population and in the three subpopulations separately (Balkan, Baranja, Medjimurje).

Star Allele	Function †	Balkan *N* (%)	Baranja *N* (%)	Medjimurje *N* (%)	Total *N* (%)
**1*	normal	50 (25.91)	84 (35.90)	78 (36.45)	212 (33.07)
2	normal	49 (25.39)	42 (17.95)	58 (27.10)	149 (23.24)
4	no function	48 (24.87)	48 (20.51)	34 (15.89)	130 (20.28)
10	decreased	6 (3.11)	12 (5.13)	21 (9.81)	39 (6.08)
22	uncertain	1 (0.52)	0	2 (0.93)	3 (0.47)
34	normal	2 (1.04)	1 (0.43)	3 (1.40)	6 (0.94)
35	normal	2 (1.04)	2 (0.85)	0	4 (0.62)
39	uncertain	3 (1.55)	1 (0.43)	5 (2.34)	9 (1.40)
41	decreased	32 (16.58)	43 (18.38)	13 (6.07)	88 (13.73)
65	uncertain	0	1 (0.43)	0	1 (0.16)
Total		193 (100)	234 (100)	214 (100)	641 (100)

† Function of star alleles was determined according to https://www.pharmvar.org/gene/CYP2D6, accessed on 18 January 2022.

Chi-square test results showed that all three Roma groups significantly differ amongst themselves in the prevalence of the five most prevalent *CYP2D6* star alleles ($\chi^2 = 34.996$, $p < 0.0001$) (Table 3). Comparing the three groups pairwise, results showed that Medjimurje Roma differ significantly from Balkan ($\chi^2 = 24.759$, $p < 0.0001$) and Baranja Roma ($\chi^2 = 22.329$, $p < 0.0001$), while Balkan and Baranja Roma do not (p = ns). We also compared the distribution of the five most frequent star alleles in the three Roma groups by comparing the prevalence of one star allele vs. other four merged ones; the distribution of four of them differs significantly among the three Roma groups (**1* $-\ \chi^2 = 6.2788$, **2* $-\ \chi^2 = 6.2276$, **10* $-\ \chi^2 = 8.6023$, **41* $-\ \chi^2 = 16.1097$; all have $p < 0.05$) (Table 3), while the distribution of **4* allele did not differ.

In addition, MJ networks were constructed to show the potential phylogenetic relationships among star alleles and their diversity in the studied Roma groups. All three MJ networks showed three clusters: one with highly predominant suballeles of **1*, the other

with suballeles of *2 and *41, and the last with suballeles of *10 and *4, suggesting that alleles *2 and *41, as well as the alleles *10 and *4, are phylogenetically close. Suballeles of *22 and *34 cluster together with suballeles of *1, suballeles of *35 cluster with suballeles of *2, while suballeles of *39 take an intermediate position between clusters in Roma groups from Baranja and Medjimurje. In Roma group from Balkans, the suballeles of *39 are placed among suballeles of *2. Suballele of *65, present in the Roma group from Baranja, is placed among suballeles of *10 (Figure 3). Six newly found haplotypes could not be translated into star alleles but according to the position in the MJ networks their classification to the star allele nomenclature could be estimated (Figure 3).

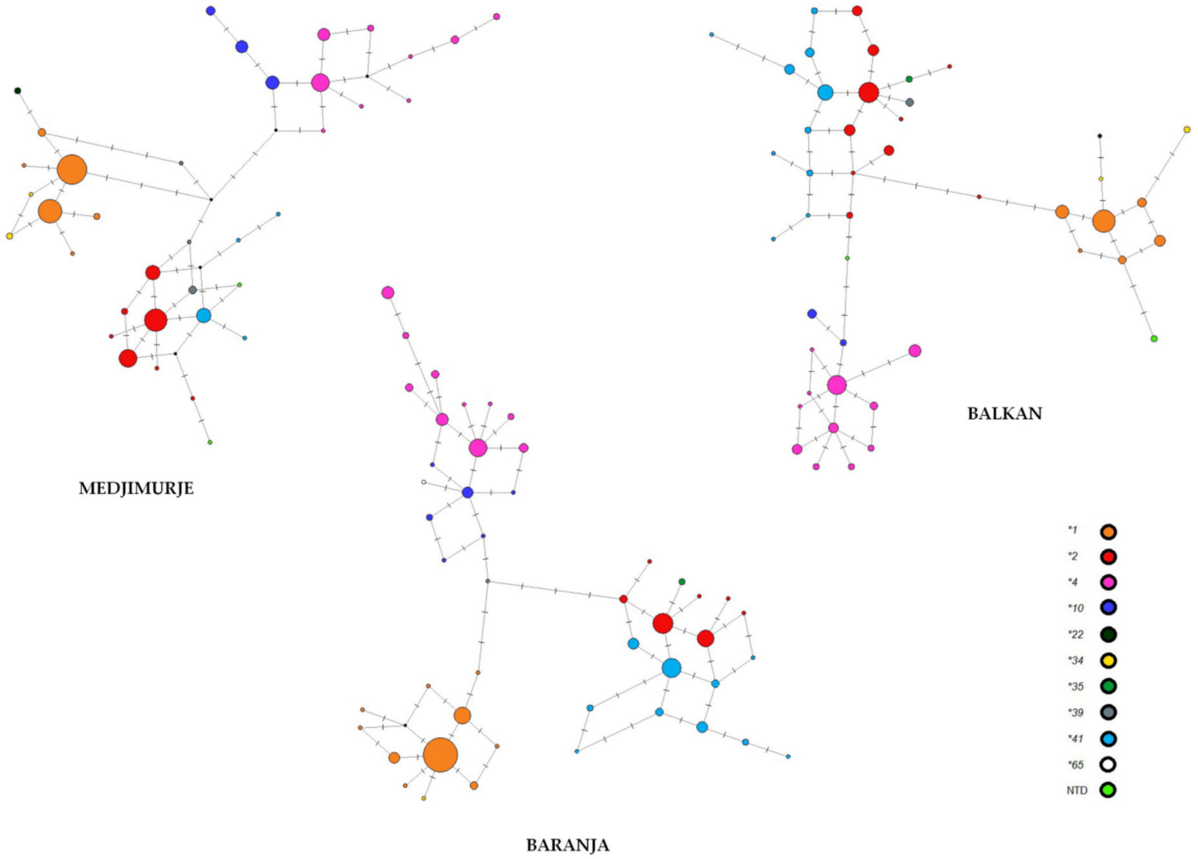


Figure 3. Median-joining networks of *CYP2D6* haplotypes in Roma from Balkan, Baranja, and Medjimurje. Haplotypes belonging to the same star allele are shown in a separate color. Distance between nodes is proportional to the numbers of SNPs whose alleles differ. SNPs are represented with lines. Size of nodes is an approximation of haplotype frequency. Black nodes represent phylogenetically possible haplotypes not found in this study. (NDT—not determined).

We also compared the frequencies of alleles *1, *2, *4, *10, *39, and *41, estimated in the Croatian Roma groups, with the population size-weighted prevalence of the same star alleles in worldwide populations, grouped according to the ethnicity (listed in Supplementary Table S2 in the Gaedigk at el. 2016 paper). Alleles *10 and *41, predominantly found in Asian populations, are present in Croatian Roma with a substantially increased prevalence compared to their European average (Figure 4). The prevalence of diplotypes and their predicted phenotypes in the three groups of Croatian Roma are shown in Table 4. We found a total of 28 diplotypes, of which 14 define normal, 13 define intermediate and one diplotype—*4/*4—defines poor metabolizers. The most prevalent diplotypes in the Croatian Roma groups are *1/*4 diplotype, which defines intermediate metabolizers, and *1/*2 diplotype, which defines normal metabolizers—these two diplotypes were found in 28.3% of the total sample. They are followed by normal metabolizing *1/*41 diplotype and intermediate metabolizing

*2/*4 diplotype prevalences. These four diplotypes were found in 46.5% of all Roma. On the other hand, 10 diplotypes were found only once in the total sample.

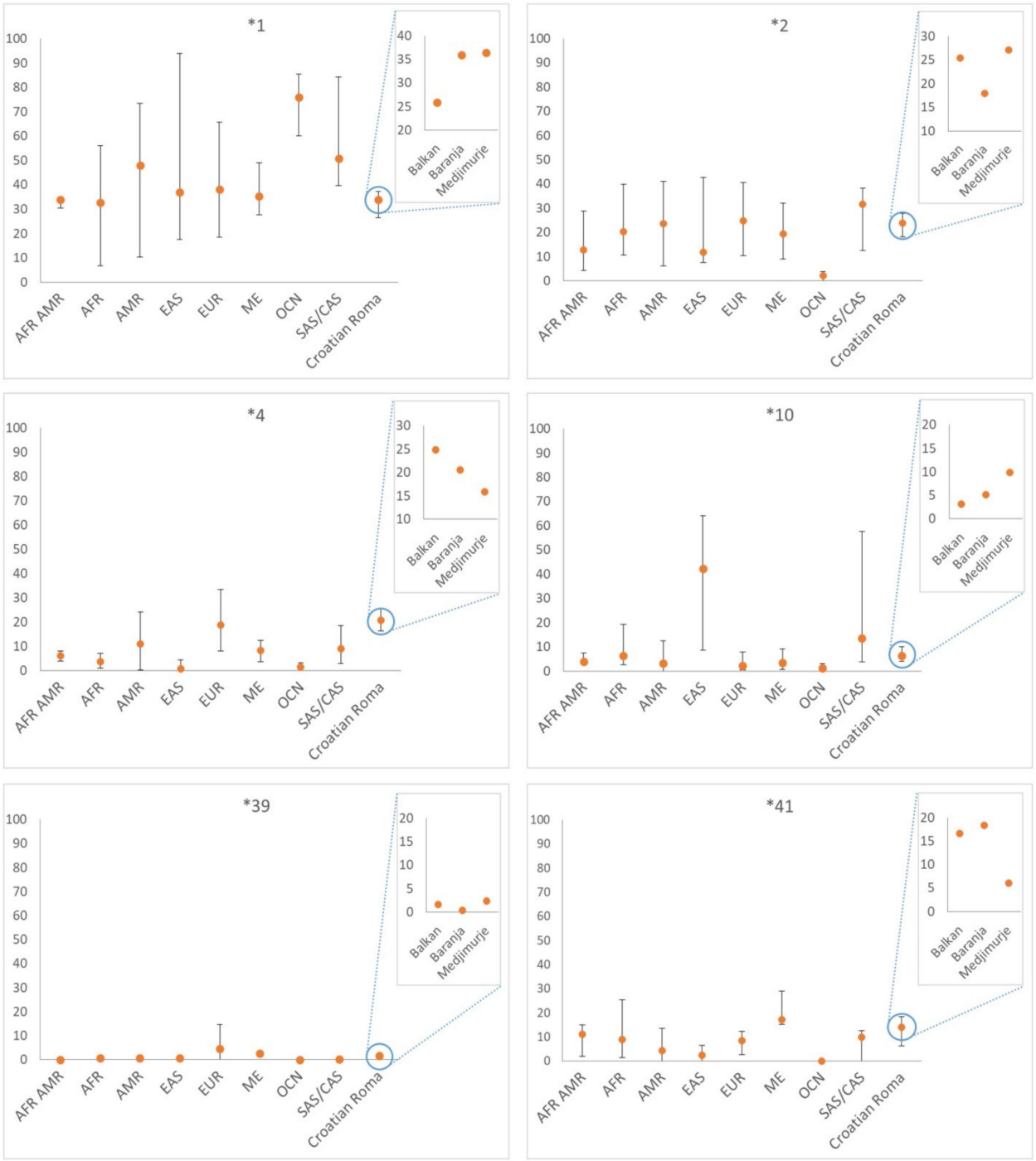


Figure 4. Distribution of *1, *2, *4, *10, *39, *41 alleles frequencies in Croatian Roma and major human populations groups (abbreviations: AFR AMR—African Americans, AFR—Africa, AMR—Americas, OCN—Oceania, EAS—East Asia, SAS/CAS—South/Central Asia, ME—Middle East, EUR—Europe).

Table 4. Distribution of star diplotypes and approximation of phenotypes in the total Croatian Roma population and in the three subpopulations separately (Balkan, Baranja, Medjimurje).

Star Diplotype	Phenotype	Balkan *N* (%)	Baranja *N* (%)	Medjimurje *N* (%)	Total *N* (%)
1/1	NM	6 (6.32)	14 (11.97)	17 (16.04)	37 (11,64)
1/2	NM	8 (8.42)	14 (11.97)	21 (19.81)	43 (13.52)
1/4	IM	16 (16.84)	22 (18.80)	9 (8.49)	47 (14.78)
1/10	NM	1 (1.05)	6 (5.13)	4 (3.77)	11 (3.46)
1/22	IM	1 (1.05)	0	1 (0.94)	2 (0.63)
1/34	NM	1 (1.05)	0	1 (0.94)	2 (0.63)
1/39	NM	1 (1.05)	0	2 (1.89)	3 (0.94)
1/41	NM	10 (10.53)	14 (11.97)	6 (5.66)	30 (9.43)
2/2	NM	10 (10.53)	4 (3.42)	8 (7.55)	22 (6.92)
2/4	IM	9 (9.47)	8 (6.84)	11 10.38)	28 (8.81)
2/10	NM	2 (2.10)	1 (0.85)	6 (5.66)	9 (2.83)
2/34	NM	0	0	1 (0.94)	1 (0.31)
2/35	NM	1 (1.05)	2 (1.71)	0	3 (0.94)
2/41	NM	8 (8.42)	9 (7.69)	3 (2.83)	20 (6.29)
4/4	PM	7 (7.37)	2 (1.71)	0	9 (2.83)
4/10	IM	2 (2.10)	5 (4.27)	9 (8.49)	16 (5.03)
4/34	IM	0	0	1 (0.94)	1 (0.31)
4/35	IM	1 (1.05)	0	0	1 (0.31)
4/39	IM	1 (1.05)	0	1 (0.94)	2 (0.63)
4/41	IM	4 (4.21)	9 (7.69)	3 (2.83)	16 (5.03)
10/10	IM	0	0	1 (0.94)	1 (0.31)
10/41	IM	1 (1.05)	0	0	1 (0.31)
22/41	IM	0	0	1 (0.94)	1 (0.31)
34/39	NM	1 (1.05)	0	0	1 (0.31)
34/41	NM	0	1 (0.85)	0	1 (0.31)
39/41	NM	0	1 (0.85)	0	1 (0.31)
41/41	IM	4 (4.21)	4 (3.42)	0	8 (2.52)
65/41	IM	0	1 (0.85)	0	1 (0.31)
Total		95 (100)	117 (100)	106 (100)	318 (100)

NM—normal metabolizer, IM—intermediate metabolizer, PM—poor metabolizer.

The most prevalent phenotype was the normal metabolizing phenotype, found in 51.6% of Balkan Roma, 56.4% of Baranja Roma, and 65.1% of Medjimurje Roma. The intermediate metabolizing phenotype was found in 41.1% of Balkan Roma, 41.9% of Baranja Roma, and 34.9% of Medjimurje Roma. The poor metabolizing phenotype was found in 7.4% of the Balkan and 1.7% of the Baranja Roma, but it was not found in any sample from the Medjimurje Roma group (Figure 5). Altogether, the normal metabolizing phenotype was found in 57.9% of examinees, the intermediate metabolizing phenotype in 39.3%, and the poor metabolizing phenotype in 2.8% of the Croatian Roma.

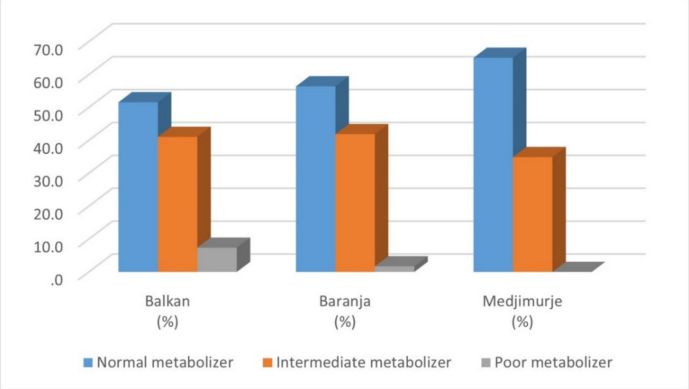


Figure 5. Distribution of estimated phenotype categories in three Roma groups.

4. Discussion

The Roma population is an example of founder populations, with centuries of socio-cultural isolation. Due to the complex history of migration combined with the cultural practice of endogamy, the Roma appear as a structured group of populations [35,38]. Indeed, studies of mitochondrial DNA (mtDNA) showed a clear divergence of the Vlax Roma from the Balkan and other Roma groups that reached Europe as part of the first migration wave [50,51]. Similar results were found with autosomal [52] and Y STR loci [53]. All current research thus points to substantial differences in the genetic make-up of diverse Roma groups [54,55]. Due to the high influence of demographic history on the gene pool of the Roma, we were interested in finding out how it reflected SNP variations within the *CYP2D6* gene.

There were 43 polymorphic SNPs within the *CYP2D6* gene in the whole sample of Croatian Roma, some of which were considered almost monomorphic in the worldwide sample, and identified only in the gnomAD database [56]. One of these SNPs is the intron variant rs368389952, found only in the Balkan Roma with a MAF of 9.18%. Its frequency reported in the gnomAD database is 0.3% in South Asian populations and in the Uygur minority it has a frequency of 1% [57]. So far it has not been reported in the PharmVar database. The second intron variant, rs566383351, is also considered almost monomorphic in the world populations, but we found this polymorphism in Baranja (13.68%), Balkan (15.82%), and Medjimurje (10.65%) Roma, which is similar to the results of Ahmed (2018) who found this SNP's minor allele in 14.59% of the Pakistani population [58]. This polymorphism has been reported in the PharmVar as a subvariant of star alleles *CYP2D6*1*, *CYP2D6*35*, and *CYP2D6*41*. It has also been reported in the Leiden Open Variation Database (LOVD) [59] as *CYP2D6*35B* with unknown effects. SNP rs374672076 is also considered monomorphic in the world population, but we have found it to be polymorphic in Balkan (8.67%) and Baranja (7.26%) Roma, which is again similar to Ahmed (2018), who found a frequency of 12.7% in the Pakistani population [58]. This SNP is an intron variant and it has been reported in the PharmVar as *CYP2D6*139.001* with unknown function. SNP rs17002852, whose MAF in our total sample is <1%, is found among Baranja Roma with a frequency of 5.98%. The highest rs17002852 MAF of 2.56% is found among Middle Eastern populations (GnomAD). This synonymous variant has been reported in the PharmVar as subvariants *CYP2D6*2.003*, *CYP2D6*2.007*, *CYP2D6*41.002*, *CYP2D6*131.001*, and *CYP2D6*149.001*. Its association with tramadol response has been reported in the ClinVar [60].

For seven SNPs (rs4987144, rs28371730, rs28371725, rs16947, rs2267447, rs3892097, and rs1065852), we found similar frequencies in the Croatian Roma, South Asian, and European populations, which is not unexpected since the Roma originated in South Asia and began their migration to Europe more than a millennium ago.

A 89 distinct *CYP2D6* haplotypes of Croatian Roma were translated into the pharmacogenetically relevant star alleles. The least number of haplotypes were found among Roma from Medjimurje, which is in line with previous findings indicating that they are the least diverse Roma group in Croatia [51,53]. In our study, the most common star allele in all three Roma groups was the *CYP2D6*1*. It is considered a reference allele and makes up most of the star alleles in the European and African populations [1,61,62]. Of the three Croatian Roma groups, the lowest frequency of the reference allele **1* was observed among Balkan Roma. As shown in Figure 4, the Balkan Roma also have the lowest frequency of **1* when compared to many world populations. Star allele *2 was the second most common allele in the Balkan and Medjimurje Roma. Balkan and Medjimurje Roma groups have a similar prevalence of *2 as the European and South Asian populations, with Medjimurje frequency being closer to European and South Asian values. This is not surprising since Naveen (2006) showed that the distribution of *CYP2D6*2* is similar between the European and South Asian populations [63]. The prevalence of *CYP2D6*2* is about 28% among Europeans, 12–29% in the Asian populations, and 16–20% in people of African ancestry [33]. Sistonen (2007) proposed that long-term selective pressure maintains a high frequency of haplotypes

encoding a fully functional enzyme, causing a homogeneous geographic distribution of *1 and *2 alleles [62].

The non-function *CYP2D6*4* allele, which is predominantly found in European populations (18%), had the highest frequency in the Balkan and Baranja Roma groups, even higher than in European populations. The prevalence of the *4 allele in the Medjimurje Roma group is lower than in the European population but still higher than in other world populations. Our results are in concordance with those for Hungarian Roma, Czech Roma, and South Asians [64,65]. *CYP2D6*4* creates a deficient protein [66] and contributes to most of the poor metabolizers observed in European populations. As a result of poor metabolization, a large accumulation of enzyme substrates occurs [67].

*CYP2D6*10* is a decreased-function allele predominantly found in East and South Asian populations, where its prevalence ranges from 9% to 44%. Its frequency in African populations is between 4–6%, among Europeans < 2% [4,33], and it is also present in the Croatian Roma (6%). According to its prevalence, the Roma from Medjimurje (10%) are closest to the South Asians, especially South Indians [63], while the Balkan Roma (3%) are closer to European populations. This allele is considered an intermediate metabolizer phenotype, and individuals homozygous for this allele are at risk for adverse events, although not as severe as in poor metabolizers [68].

Finally, a decreased-function allele *CYP2D6*41* is found in Croatian Roma at 14% frequency. Its prevalence among African populations is 4–11.5%, in Asian populations 2–12%, and about 9% in European populations [33]. Roma from Medjimurje (6%) are closer to European and South Asian populations, while the Balkan (17%) and Baranja Roma (18%) are closer to Middle East populations, which show the highest frequencies of *41 allele in the world (Figure 4).

Distribution of the five most frequent star alleles (*1, *2, *4, *10, and *41), which accounted for over 95% of the variance in all three Croatian Roma groups, did not significantly differ between Roma from Balkan and Baranja, while Roma from Medjimurje significantly differed from both these groups. The results on Roma from Medjimurje are in line with previous findings that showed the highest level of isolation compared to other Roma groups [51]. Although this research is missing the determination of structural variants, such as gene duplications, which are important for the accurate determination of phenotypes [37], we assessed the metabolizing phenotype from diplotype data. Distributions of metabolizing phenotypes are similar among Roma groups, but Roma from Medjimurje, with the highest frequency of normal metabolizers, the lowest frequency of intermediate metabolizers, and none of the poor metabolizers, are the most distinct. The results of Medjimurje Roma are similar to Hungarian Roma, which also had no poor metabolizers [64].

Because of their socio-economic status, the Roma have less access to medical care and are at higher risk of diseases like diabetes, cardiovascular diseases, and other complex diseases [69]. Since CYP2D6 metabolizes many commonly used drugs [70], it is of great interest for research not only in the general population but also in isolated or minority populations. Studies on clinical effects of several antiarrhythmic drugs including metoprolol, timolol and propafenone, and antidepressants and antipsychotics have not been unanimous. It is assumed that poor/intermediate metabolizers are prone to adverse drug reactions. Furthermore, for antidepressants and antipsychotics, there is a risk of overexposure in poor/intermediate metabolizers and underexposure in normal metabolizers [68]. For several opioid drugs (codeine, oxycodone, and tramadol) used to treat pain, genotypes have been shown to affect their efficacy and safety [68]. Cancer research studies are not in agreement with the role of CYP2D6 in the development of cancer [71,72]. Still, this enzyme is involved in the metabolism of cytotoxic drugs such as tamoxifen, and it has been shown that both poor and ultrarapid CYP2D6 metabolizers of tamoxifen have a worse prognosis compared with normal metabolizers [73,74].

Pharmacogenomic research is an important tool in drug development and health system improvement that leads to personalized medicine. In populations that are un-

likely to have access to personalized medicine, a population profiling like this one can be of interest for medical practitioners since the results of such research can provide the basis for the avoidance (or careful monitoring of the effects) of the administration of drugs that contain substances that are not properly metabolized among the members of that population.

5. Conclusions

Summarizing our results, we can say that demographic history, predominantly migrations (from India to Balkans and across Southeast European areas) and endogamy, has indeed influenced the distribution of variations within the *CYP2D6* gene. It can be seen in the accumulation of globally rare variants which is the result of genetic drift that operates in isolated populations such as the Roma. Additionally, traces of their South Asian origin can be seen in the frequencies of polymorphic variants that are similar to Asian populations in many SNPs, as well as in elevated frequencies of star alleles **10* and **41*. Given metabolizing phenotype estimates, Croatian Roma generally have low levels of poor metabolizers. The three socio-culturally different Roma groups studied differ significantly in the distribution of star alleles, which confirms the importance of studying different Roma groups separately.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm12030374/s1, Table S1: Major and minor allele frequencies of polymorphic loci in East Asian, South Asian, European, and Croatian Roma populations; Table S2: CYP2D6 haplotype list; Table S3: Worldwide frequencies of most frequent star alleles in Roma population.

Author Contributions: Conceptualization, M.P.S. and M.Z.P.; methodology, A.S.M., M.P.S.; validation, all; formal analysis, A.S.M., M.Z.P.; investigation, Ž.T., B.P.; A.S.M.; data curation, T.Š.-J., Ž.T.; writing—original draft preparation, A.S.M., M.Z.P.; writing—review and editing, All; visualization, M.Š.; supervision, M.P.S.; funding acquisition, M.P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Croatian Science Foundation (IP-2014-09-4454 and DOK-2018-01-4817 to M.P.S.).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Scientific Committee and the Ethics Committee of the Institute for Anthropological Research (RN 1.14-1611/14).

Informed Consent Statement: All Roma participated in the study voluntarily and were informed about the goals, methods, and expectations of the study with the help of linguistically and culturally competent and trained Roma volunteers, after which they gave their informed consent.

Data Availability Statement: All data analyzed in this study are available at http://roma.inantro.hr/en/, accessed on 18 January 2022. In case of using this database for further analyses, please cite this publication. If further clarification is required, contact the corresponding author.

Acknowledgments: We are deeply grateful to the Roma people for their kindness and interest in participation in this study. We would like to thank the anonymous reviewers for their suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J. The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am. J. Hum. Genet.* **1989**, *45*, 889–904. [PubMed]
2. Heim, M.H.; Meyer, U.A. Evolution of a highly polymorphic human cytochrome P450 gene cluster: CYP2D6. *Genomics* **1992**, *14*, 49–58. [CrossRef]
3. Daly, A.K.; Brockmöller, J.; Broly, F.; Eichelbaum, M.; Evans, W.E.; Gonzalez, F.J.; Huang, J.D.; Idle, J.R.; Ingelman-Sundberg, M.; Ishizaki, T.; et al. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics* **1996**, *6*, 193–201. [CrossRef] [PubMed]

4. Gaedigk, A.; Sangkuhl, K.; Whirl-Carrillo, M.; Klein, T.; Leeder, J.S. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* **2017**, *19*, 69–76. [CrossRef]
5. Hicks, J.K.; Swen, J.J.; Gaedigk, A. Challenges in CYP2D6 phenotype assignment from genotype data: A critical assessment and call for standardization. *Curr. Drug Metab.* **2014**, *15*, 218–232. [CrossRef]
6. Caudle, K.E.; Sangkuhl, K.; Whirl-Carrillo, M.; Swen, J.J.; Haidar, C.E.; Klein, T.E.; Gammal, R.S.; Relling, M.V.; Scott, S.A.; Hertz, D.L.; et al. Standardizing CYP2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clin. Transl. Sci.* **2019**, *13*, 116–124. [CrossRef]
7. Human Cytochrome P450 (CYP) Allele Nomenclature Database. Available online: http://www.cypalleles.ki.se/ (accessed on 15 August 2021).
8. PHARMGKB. Available online: https://www.pharmgkb.org/ (accessed on 15 August 2021).
9. He, Z.X.; Chen, X.W.; Zhou, Z.W.; Zhou, S.F. Impact of physiological, pathological and environmental factors on the expression and activity of human cytochrome P450 2D6 and implications in precision medicine. *Drug Metab. Rev.* **2015**, *47*, 470–519. [CrossRef]
10. Ingelman-Sundberg, M.; Sim, S.C.; Gomez, A.; Rodriguez-Antona, C. Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.* **2007**, *116*, 496–526. [CrossRef]
11. Zanger, U.M.; Schwab, M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* **2013**, *138*, 103–141. [CrossRef]
12. Williams, I.S.; Gatchie, L.; Bharate, S.B.; Chaudhuri, B. Biotransformation, Using Recombinant CYP450-Expressing Baker's Yeast Cells, Identifies a Novel CYP2D6.10A122V Variant Which Is a Superior Metabolizer of Codeine to Morphine Than the Wild-Type Enzyme. *ACS Omega* **2018**, *3*, 8903–8912. [CrossRef]
13. Zanger, U.M.; Turpeinen, M.; Klein, K.; Schwab, M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal. Bioanal. Chem.* **2008**, *392*, 1093–1108. [CrossRef]
14. Fleeman, N.; Dundar, Y.; Dickson, R.; Jorgensen, A.; Pushpakom, S.; McLeod, C.; Pirmohamed, M.; Walley, T. Cytochrome P450 testing for prescribing antipsychotics in adults with schizophrenia: Systematic review and meta-analyses. *Pharm. J.* **2010**, *11*, 1–14. [CrossRef]
15. Stingl, J.C.; Brockmöller, J.; Viviani, R. Genetic variability of drug-metabolizing enzymes: The dual impact on psychiatric therapy and regulation of brain function. *Mol. Psychiatry* **2012**, *18*, 273–287. [CrossRef]
16. Gaedigk, A. Complexities of CYP2D6gene analysis and interpretation. *Int. Rev. Psychiatry* **2013**, *25*, 534–553. [CrossRef]
17. Hicks, J.K.; Swen, J.J.; Thorn, C.F.; Sangkuhl, K.; Kharasch, E.D.; Ellingrod, V.L.; Skaar, T.C.; Müller, D.J.; Gaedigk, A.; Stingl, J.C. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clin. Pharmacol. Ther.* **2015**, *98*, 127–134. [CrossRef]
18. Beoris, M.; Amos Wilson, J.; Garces, J.A.; Lukowiak, A.A. CYP2D6 copy number distribution in the US population. *Pharm. Genom.* **2016**, *26*, 96–99. [CrossRef]
19. Christensen, P.M.; Gotzsche, P.C.; Brosen, K. The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of Parkinson's disease: A meta-analysis. *Pharmacogenetics* **1998**, *8*, 473–479. [CrossRef]
20. Lu, Y.; Peng, Q.; Zeng, Z.; Wang, J.; Deng, Y.; Xie, L.; Mo, C.; Zeng, J.; Qin, X.; Li, S. CYP2D6 phenotypes and Parkinson's disease risk: A meta-analysis. *J. Neurol. Sci.* **2014**, *336*, 161–168. [CrossRef]
21. Ur Rasheed, M.S.; Mishra, A.K.; Singh, M.P. Cytochrome P450 2D6 and Parkinson's Disease: Polymorphism, Metabolic Role, Risk and Protection. *Neurochem. Res.* **2017**, *42*, 3353–3361. [CrossRef]
22. Patsopoulos, N.A.; Ntzani, E.E.; Zintzaras, E.; Ioannidis, J.P. CYP2D6 polymorphisms and the risk of tardive dyskinesia in schizophrenia: A meta-analysis. *Pharm. Genom.* **2005**, *15*, 151–158. [CrossRef]
23. Scordo, M.G.; Dahl, M.L.; Spina, E.; Cordici, F.; Arena, M.G. No association between CYP2D6 polymorphism and Alzheimer's disease in an Italian population. *Pharmacol. Res.* **2006**, *53*, 162–165. [CrossRef]
24. Ma, S.L.; Tang, N.L.S.; Wat, K.H.Y.; Tang, J.H.Y.; Lau, K.H.; Law, C.B.; Chiu, J.; Tam, C.C.W.; Poon, T.K.; Lin, K.L.; et al. Effect of CYP2D6 and CYP3A4 Genotypes on the Efficacy of Cholinesterase Inhibitors in Southern Chinese Patients with Alzheimer's Disease. *Am. J. Alzheimers Dis. Other Dement.* **2019**, *34*, 302–307. [CrossRef]
25. Agundez, J. Cytochrome P450 Gene Polymorphism and Cancer. *Curr. Drug Metab.* **2004**, *5*, 211–224. [CrossRef]
26. Rodriguez-Antona, C.; Gomez, A.; Karlgren, M.; Sim, S.C.; Ingelman-Sundberg, M. Molecular genetics and epigenetics of the cytochrome P450 gene family and its relevance for cancer risk and treatment. *Hum. Genet.* **2009**, *127*, 1–17. [CrossRef]
27. Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. *Pharm. J.* **2004**, *5*, 6–13. [CrossRef]
28. Aklillu, E.; Herrlin, K.; Gustafsson, L.L.; Bertilsson, L.; Ingelman-Sundberg, M. Evidence for environmental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. *Pharmacogenetics* **2002**, *12*, 375–383. [CrossRef]
29. Podgorná, E.; Diallo, I.; Vangenot, C.; Sanchez-Mazas, A.; Sabbagh, A.; Černý, V.; Poloni, E.S. Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. *BMC Evol. Biol.* **2015**, *15*, 263. [CrossRef]

30. Fuselli, S.; de Filippo, C.; Mona, S.; Sistonen, J.; Fariselli, P.; Destro-Bisol, G.; Barbujani, G.; Bertorelle, G.; Sajantila, A. Evolution of detoxifying systems: The role of environment and population history in shaping genetic diversity at human CYP2D6 locus. *Pharmacogenet. Genom.* **2010**, *20*, 485–499. [CrossRef]
31. Li, J.; Zhang, L.; Zhou, H.; Stoneking, M.; Tang, K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* **2011**, *20*, 528–540. [CrossRef]
32. LLerena, A.; Naranjo, M.E.; Rodrigues-Soares, F.; Penas-LLedó, E.M.; Fariñas, H.; Tarazona-Santos, E. Interethnic variability ofCYP2D6alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin. Drug Metab. Toxicol.* **2014**, *10*, 1569–1583. [CrossRef]
33. Pratt, V.M.; Cavallari, L.H.; Del Tredici, A.L.; Gaedigk, A.; Hachad, H.; Ji, Y.; Kalman, L.V.; Ly, R.C.; Moyer, A.M.; Scott, S.A.; et al. Recommendations for Clinical CYP2D6 Genotyping Allele Selection. *J. Mol. Diagn.* **2021**, *23*, 1047–1064. [CrossRef] [PubMed]
34. Zhang, H.; De, T.; Zhong, Y.; Perera, M.A. The Advantages and Challenges of Diversity in Pharmacogenomics: Can Minority Populations Bring Us Closer to Implementation? *Clin. Pharmacol. Ther.* **2019**, *106*, 338–349. [CrossRef] [PubMed]
35. Fraser, A. *The Gypsies*; Blackwell Publishers: Oxford, UK, 1992.
36. Hancock, I.F. *We Are the Romani People*; University of Herdforshire Press: Hatfield, UK, 2002.
37. Gresham, D.; Morar, B.; Underhill, P.A.; Passarino, G.; Lin, A.A.; Wise, C.; Angelicheva, D.; Calafell, F.; Oefner, P.J.; Shen, P.; et al. Origins and divergence of the Roma (gypsies). *Am. J. Hum. Genet.* **2001**, *69*, 1314–1331. [CrossRef] [PubMed]
38. Chaix, R.; Austerlitz, F.; Morar, B.; Kalaydjieva, L.; Heyer, E. Vlax Roma history: What do coalescent-based methods tell us? *Eur. J. Hum. Genet.* **2004**, *12*, 285–292. [CrossRef]
39. Škarić-Jurić, T.; Tomas, Ž.; Zajc Petranović, M.; Božina, N.; Smolej Narančić, N.; Janićijević, B.; Salihović, M.P. Characterization of ADME genes variation in Roma and 20 populations worldwide. *PLoS ONE* **2018**, *13*, e0207671. [CrossRef]
40. Miller, S.A.; Dykes, D.D.; Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **1988**, *16*, 1215. [CrossRef]
41. Campbell, N.R.; Harmon, S.A.; Narum, S.R. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol. Ecol. Resour.* **2015**, *15*, 855–867. [CrossRef]
42. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
43. Stephens, M.; Smith, N.J.; Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **2001**, *68*, 978–989. [CrossRef]
44. Stephens, M.; Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **2003**, *73*, 1162–1169. [CrossRef]
45. Pharmacogene Variation (PharmVar) Consortium—CYP2D6. Available online: https://www.pharmvar.org/gene/CYP2D6 (accessed on 15 August 2021).
46. Clinical Pharmacogenetics Implementation Consortium—Term Standardization. Available online: https://cpicpgx.org/resources/term-standardization (accessed on 15 August 2021).
47. Excoffier, L.; Lischer, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [CrossRef]
48. Bandelt, H.J.; Forster, P.; Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [CrossRef]
49. Free Phylogenetic Network Software. Available online: https://www.fluxus-engineering.com/sharenet.htm (accessed on 10 January 2021).
50. Mendizabal, I.; Valente, C.; Gusmão, A.; Alves, C.; Gomes, V.; Goios, A.; Parson, W.; Calafell, F.; Alvarez, L.; Amorim, A.; et al. Reconstructing the Indian origin and dispersal of the European Roma: A maternal genetic perspective. *PLoS ONE* **2011**, *6*, e15988. [CrossRef]
51. Salihović, M.P.; Barešić, A.; Klarić, I.M.; Cukrov, S.; Lauc, L.B.; Janićijević, B. The role of the Vlax Roma in shaping the European Romani maternal genetic history. *Am. J. Phys. Anthropol.* **2011**, *146*, 262–270. [CrossRef]
52. Gusmão, A.; Valente, C.; Gomes, V.; Alves, C.; Amorim, A.; Prata, M.J.; Gusmão, L. A genetic historical sketch of European Gypsies: The perspective from autosomal markers. *Am. J. Phys. Anthropol.* **2010**, *141*, 507–514. [CrossRef]
53. Klarić, I.M.; Salihović, M.P.; Lauc, L.B.; Zhivotovsky, L.A.; Rootsi, S.; Janićijević, B. Dissecting the molecular architecture and origin of Bayash Romani patrilineages: Genetic influences from South-Asia and the Balkans. *Am. J. Phys. Anthropol.* **2009**, *138*, 333–342. [CrossRef]
54. Bianco, E.; Laval, G.; Font-Porterias, N.; García-Fernández, C.; Dobon, B.; Sabido-Vera, R.; Sukarova Stefanovska, E.; Kučinskas, V.; Makukh, H.; Pamjav, H.; et al. Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes. *Mol. Biol. Evol.* **2020**, *37*, 3175–3187. [CrossRef]
55. Mendizabal, I.; Lao, O.; Marigorta, U.M.; Wollstein, A.; Gusmão, L.; Ferak, V.; Ioana, M.; Jordanova, A.; Kaneva, R.; Kouvatsi, A.; et al. Reconstructing the population history of European Romani from genome-wide data. *Curr. Biol.* **2012**, *22*, 2342–2349. [CrossRef]
56. Genome Aggregation Database. Available online: https://gnomad.broadinstitute.org/ (accessed on 17 June 2021).
57. He, X.; He, N.; Ren, L.; Ouyang, Y.; Zhang, N.; Ma, Y.; Yuan, D.; Kang, L.; Jin, T. Genetic polymorphisms analysis of CYP2D6 in the Uygur population. *BMC Genom.* **2016**, *17*, 409. [CrossRef]

58. Ahmed, S.; Zhou, J.; Zhou, Z.; Chen, S.Q. Genetic Polymorphisms and In Silico Mutagenesis Analyses of CYP2C9, CYP2D6, and CYPOR Genes in the Pakistani Population. *Genes* **2018**, *9*, 514. [CrossRef]
59. Leiden Open Variation Database. Available online: https://databases.lovd.nl (accessed on 15 August 2021).
60. ClinVar—RCV001028788.2. Available online: https://www.ncbi.nlm.nih.gov/clinvar/RCV001028788 (accessed on 15 August 2021).
61. Bradford, L.D. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. *Pharmacogenomics* **2002**, *3*, 229–243. [CrossRef]
62. Sistonen, J.; Sajantila, A.; Lao, O.; Corander, J.; Barbujani, G.; Fuselli, S. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenet. Genom.* **2007**, *17*, 93–101. [CrossRef]
63. Naveen, A.T.; Adithan, C.; Soya, S.S.; Gerard, N.; Krishnamoorthy, R. CYP2D6 Genetic Polymorphism in South Indian Populations. *Biol. Pharm. Bull.* **2006**, *29*, 1655–1658. [CrossRef]
64. Weber, A.; Szalai, R.; Sipeky, C.; Magyari, L.; Melegh, M.; Jaromi, L.; Matyas, P.; Duga, B.; Kovesdi, E.; Hadzsiev, K.; et al. Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples. *Pharmacol. Rep.* **2015**, *67*, 460–464. [CrossRef] [PubMed]
65. Dlouhá, L.; Adámková, V.; Šedová, L.; Olišarová, V.; Hubáček, J.A.; Tóthová, V. Five genetic polymorphisms of cytochrome P450 enzymes in the Czech non-Roma and Czech Roma population samples. *Drug Metab. Pers. Ther.* **2020**, *35*, 20200103. [CrossRef] [PubMed]
66. Gough, A.C.; Miles, J.S.; Spurr, N.K.; Moss, J.E.; Gaedigk, A.; Eichelbaum, M.; Wolf, C.R. Identification of the primary gene defect at the cytochrome P450 CYP2D locus. *Nature* **1990**, *347*, 773–776. [CrossRef] [PubMed]
67. Gomes, L.; Lemos, M.C.; Paiva, I.; Ribeiro, C.; Carvalheiro, M.; Regateiro, F.J. CYP2D6 genetic polymorphisms are associated with susceptibility to pituitary tumors. *Acta Med. Port.* **2005**, *18*, 339–343. [PubMed]
68. Zanger, U.M.; Raimundo, S.; Eichelbaum, M. Cytochrome P450 2D6: Overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2004**, *369*, 23–37. [CrossRef]
69. Petranović, M.Z.; Rizzieri, A.E.; Sivaraj, D.; Narančić, N.S.; Škarić-Jurić, T.; Celinšćak, Ž.; Marković, A.S.; Salihović, M.P.; Kalászi, J.; Kalászi, M.; et al. CVD Risk Factors in the Ukrainian Roma and Meta-Analysis of Their Prevalence in Roma Populations Worldwide. *J. Pers. Med.* **2021**, *11*, 1138. [CrossRef]
70. Gardiner, S.J.; Begg, E.J. Pharmacogenetics, Drug-Metabolizing Enzymes, and Clinical Practice. *Pharmacol. Rev.* **2006**, *58*, 521–590. [CrossRef]
71. Christensen, P.M.; Gøtzsche, P.C.; Brøsen, K. The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of lung cancer: A meta-analysis. *Eur. J. Clin. Pharmacol.* **1997**, *51*, 389–393. [CrossRef]
72. Wolf, C.R.; Smith, C.A.; Bishop, T.; Forman, D.; Gough, A.C.; Spurr, N.K. CYP 2D6 genotyping and the association with lung cancer susceptibility. *Pharmacogenetics* **1994**, *4*, 104–106. [CrossRef]
73. He, W.; Grassmann, F.; Eriksson, M.; Eliasson, E.; Margolin, S.; Thorén, L.; Hall, P.; Czene, K. CYP2D6 Genotype Predicts Tamoxifen Discontinuation and Prognosis in Patients with Breast Cancer. *J. Clin. Oncol.* **2020**, *38*, 548–557. [CrossRef]
74. He, W.; Eriksson, M.; Eliasson, E.; Grassmann, F.; Bäcklund, M.; Gabrielson, M.; Hammarström, M.; Margolin, S.; Thorén, L.; Wengström, Y.; et al. CYP2D6 genotype predicts tamoxifen discontinuation and drug response: A secondary analysis of the KARISMA trial. *Ann. Oncol.* **2021**, *32*, 1286–1293. [CrossRef]

Supp. Table 1 Major and minor allele frequencies of polymorphic loci in East Asian, South Asian, European and Croatian Roma populations

										CROATIAN ROMA					
Position at Chr22 (GRCh38.p13)	dbSNP ID	REF	ALT	EAST ASIAN minor	EAST ASIAN major	SOUTH ASIAN minor	SOUTH ASIAN major	EUROPEAN* minor	EUROPEAN* major	Balkan minor	Balkan major	Medjimurje minor	Medjimurje major	Baranja minor	Baranja major
42126611	rs1135840	G	C	0,6895	0,3105	0,5517	0,4483	0,5588	0,4412	0,7194	0,2806	0,6121	0,3879	0,6379	0,3621
42126963	rs28371732	G	A	0,0000	1,0000	0,0021	0,9979	0,0072	0,9928	0,0000	1,0000	0,0048	0,9952	0,0000	1,0000
42127001	rs4987144	C	T	0,1400	0,8600	0,3725	0,6275	0,3263	0,6737	0,3265	0,6735	0,2870	0,7130	0,3162	0,6838
42127207	rs28371730	G	A	0,1463	0,8537	0,3806	0,6194	0,3320	0,6680	0,3673	0,6327	0,3519	0,6481	0,3675	0,6325
42127209	rs2004511	A	G	0,5542	0,4458	0,1714	0,8286	0,2268	0,7732	0,3673	0,6327	0,4028	0,5972	0,3376	0,6624
42127246	rs1269631565	T	TT	0,0000	1,0000	0,0010	0,9990	0,0003	0,9997	0,0204	0,9796	0,1019	0,8981	0,0470	0,9530
42127307	rs867985262	G	A	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0046	0,9954	0,0000	1,0000
42127313	rs79596243	A	G	0,0000	1,0000	0,0011	0,9989	0,0004	0,9996	0,0000	1,0000	0,0046	0,9954	0,0000	1,0000
42127356	rs28371729	C	A	0,0002	0,9998	0,0038	0,9962	0,0299	0,9701	0,0102	0,9898	0,0046	0,9954	0,0043	0,9957
42127398	rs28578778	T	C	0,0000	1,0000	0,0142	0,9858	0,0111	0,9889	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42127407	rs1985842	A	C	0,7039	0,2961	0,5537	0,4463	0,5599	0,4401	0,7143	0,2857	0,6157	0,3843	0,6410	0,3590
42127612	rs200335621	C	T	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0102	0,9898	0,0093	0,9907	0,0427	0,9573
42127631	rs141009491	G	C	0,0000	1,0000	0,0000	1,0000	0,0004	0,9996	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42127803	rs28371725	G	A	0,0330	0,9670	0,1261	0,8739	0,0955	0,9045	0,1648	0,8352	0,0728	0,9272	0,1858	0,8142
42127941	rs16947	C	T	0,1490	0,8510	0,3807	0,6193	0,3347	0,6653	0,4278	0,5722	0,3252	0,6748	0,3784	0,6216
128177-421281	rs762158210	GAGAA	GA	0,0000	1,0000	0,0002	0,9998	0,0000	1,0000	0,0145	0,9855	0,0000	1,0000	0,0000	1,0000
42128242	rs35742686	CAG	CG	0,0000	1,0000	0,0017	0,9983	0,0170	0,9830	0,0054	0,9946	0,0000	1,0000	0,0000	1,0000
42128308	rs28371717	G	T	0,0000	1,0000	0,0042	0,9958	0,0129	0,9871	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42128321	rs17002852	T	C	0,0002	0,9998	0,0137	0,9863	0,0079	0,9921	0,0357	0,9643	0,0000	1,0000	0,0598	0,9402
42128694	rs2267447	A	G	0,5514	0,4486	0,1701	0,8299	0,2222	0,7778	0,2755	0,7245	0,2268	0,7732	0,2521	0,7479
42128945	rs3892097	G	A	0,0068	0,9932	0,1081	0,8919	0,1986	0,8014	0,2449	0,7551	0,1574	0,8426	0,2051	0,7949
42129042	rs1135824	A	G	0,0000	1,0000	0,0006	0,9994	0,0002	0,9998	0,0306	0,9694	0,0000	1,0000	0,0000	1,0000
42129130	rs1058164	G	C	0,6898	0,3102	0,5542	0,4458	0,5525	0,4475	0,7143	0,2857	0,6111	0,3889	0,6410	0,3590
42129351	rs376056664	T	C	0,0000	1,0000	0,0000	1,0000	0,0004	0,9996	0,0052	0,9948	0,0000	1,0000	0,0000	1,0000
42129498	rs189736703	G	A	0,0000	1,0000	0,0012	0,9988	0,0039	0,9961	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42129623	rs1081004	G	A	0,0000	1,0000	0,0113	0,9888	0,0709	0,9291	0,0256	0,9744	0,0370	0,9630	0,0513	0,9487
42129722	rs368389952	G	A	0,0000	1,0000	0,0029	0,9971	0,0003	0,9997	0,0918	0,9082	0,0000	1,0000	0,0000	1,0000
42129754	rs1081003	C	T	0,5368	0,4632	0,0780	0,9220	0,0254	0,9746	0,0204	0,9796	0,0556	0,9444	0,0129	0,9871
42129796	rs28371705	C	G	0,0041	0,9959	0,0889	0,9111	0,1945	0,8055	0,0357	0,9643	0,0000	1,0000	0,0086	0,9914
42129809	rs28371704	A	G	0,0029	0,9971	0,0866	0,9134	0,1925	0,8075	0,0309	0,9691	0,0000	1,0000	0,0086	0,9914
42129819	rs28371703	C	A	0,0029	0,9971	0,0862	0,9138	0,1918	0,8082	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42130522	rs29001678	C	T	0,0012	0,9988	0,0192	0,9808	0,0342	0,9658	0,0122	0,9878	0,0103	0,9897	0,0297	0,9703
42130547	rs1081000	A	G	0,0201	0,9799	0,0938	0,9062	0,0866	0,9134	0,0567	0,9433	0,0000	1,0000	0,0085	0,9915
42130692	rs1065852	C	T	0,5413	0,4587	0,1681	0,8319	0,2225	0,7775	0,2806	0,7194	0,2315	0,7685	0,2607	0,7393
42130710	rs138100349	C	T	0,0002	0,9998	0,0002	0,9998	0,0029	0,9971	0,0051	0,9949	0,0093	0,9907	0,0128	0,9872
42130761	rs769258	G	A	0,0027	0,9973	0,0075	0,9925	0,0611	0,9389	0,0153	0,9847	0,0000	1,0000	0,0085	0,9915
42130993	rs372204775	G	A	0,0002	0,9998	0,0251	0,9749	0,0005	0,9995	0,0103	0,9897	0,0417	0,9583	0,0256	0,9744
42131145	rs530422334	A	G	0,0002	0,9998	0,0006	0,9994	0,0003	0,9997	0,0000	1,0000	0,0324	0,9676	0,0171	0,9829
42131156	rs1080992	G	A	0,0002	0,9998	0,0015	0,9985	0,0135	0,9865	0,0000	1,0000	0,0093	0,9907	0,0000	1,0000
42131188	rs769811346	G	T	0,0000	1,0000	0,0000	1,0000	0,0003	0,9997	0,0000	1,0000	0,0000	1,0000	0,0043	0,9957
42131189	rs374672076	G	C	0,0010	0,9990	0,0004	0,9996	0,0009	0,9991	0,0867	0,9133	0,0463	0,9537	0,0726	0,9274
42131222	rs566383351	C	T	0,0000	1,0000	0,0002	0,9998	0,0004	0,9996	0,1582	0,8418	0,1065	0,8935	0,1368	0,8632

Source of allele frequncy data: gnomAD (https://gnomad.broadinstitute.org)

* allele frequencies without Finish population

Supp. Table 2 *CYP2D6* haplotype list

		dbSNP ID	rs1135840	rs4987144	rs28371730	rs2004511	rs28371729	rs1985842	rs200335621	rs28371725	rs16947	rs17002852	rs2267447	rs3892097	rs1135824	rs1058164	rs1081004
Haplotype ID	star alellel	HGVS name	9200G>C	8810C>T	8604G>A	8602A>G	8455C>A	8404A>C	8199C>T	8008G>A	7870C>T	7490T>C	7117A>G	6866G>A	6769A>G	6681G>C	6188G>A
Hap1	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap2	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap3	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap4	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	**C**	G
Hap5	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap6	*1		G	C	G	A	C	A	C	G	c	T	A	G	A	G	G
Hap7	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	**A**
Hap8	*1		G	C	G	**G**	C	**C**	C	G	C	T	A	G	A	G	G
Hap9	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap10	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	G
Hap11	*1		G	C	G	A	C	A	C	G	C	T	A	G	A	G	**A**
Hap12	*1		G	C	G	**G**	C	A	C	G	C	T	A	G	A	G	G
Hap13	*1		G	C	G	**G**	C	A	C	G	C	T	A	G	A	G	G
Hap14	*1		G	C	G	**G**	C	A	C	G	C	T	A	G	A	G	G
Hap15	*2		**C**	C	G	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap16	*2		**C**	C	G	**G**	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap17	*2		**C**	C	G	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap18	*2		**C**	**T**	**A**	**G**	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap19	*2		**C**	C	G	A	C	A	C	G	**T**	T	A	G	A	G	G
Hap20	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap21	*2		**C**	C	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap22	*2		**C**	**T**	G	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap23	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap24	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap25	*2		**C**	C	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap26	*2		**C**	C	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap27	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap28	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap29	*2		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	G	G

Hap30	*2		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap31	*2		**C**	**T**	**A**	A	**A**	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap32	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap33	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap34	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	**G**	**C**	G
Hap35	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	**G**	**C**	G
Hap36	*4		**C**	C	G	**G**	**A**	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap37	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap38	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap39	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	**A**
Hap40	*4		**C**	C	G	**G**	C	**C**	C	G	C	**C**	**G**	**A**	A	**C**	G
Hap41	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap42	*4		**C**	C	G	**G**	C	**C**	**T**	G	C	T	**G**	**A**	A	**C**	G
Hap43	*4		**C**	C	G	**G**	**A**	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap44	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap45	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap46	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap47	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap48	*4		**C**	C	G	**G**	C	**C**	**T**	G	C	T	A	**A**	A	**C**	G
Hap49	*4		**C**	C	G	A	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap50	*4		G	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap51	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	A	**A**	A	**C**	G
Hap52	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	A	**A**	A	**C**	G
Hap53	*4		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	**A**	A	**C**	G
Hap54	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	A	G	A	**C**	G
Hap55	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	A	G	A	**C**	G
Hap56	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	G	A	**C**	G
Hap57	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	G	A	**C**	G
Hap58	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	G	A	**C**	G
Hap59	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	G	A	**C**	G
Hap60	*10		**C**	C	G	**G**	C	**C**	C	G	C	T	**G**	G	A	**C**	G
Hap61	*22		G	C	G	A	C	A	C	G	C	T	A	G	A	G	**A**
Hap62	*34		G	C	G	**G**	C	A	C	G	**T**	T	A	G	A	G	G

Hap63	*34		G	C	G	**G**	C	A	C	G	**T**	T	A	G	A	G	G
Hap64	*34		G	C	G	A	C	A	C	G	**T**	T	A	G	A	G	G
Hap65	*35		**C**	**T**	**A**	A	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap66	*39		**C**	C	G	**G**	C	**C**	C	G	C	T	A	G	A	**C**	G
Hap67	*39		**C**	C	G	A	C	**C**	C	G	C	T	A	G	A	**C**	**A**
Hap68	*39		**C**	C	**A**	A	C	**C**	C	G	C	T	A	G	A	**C**	G
Hap69	*39		**C**	**T**	**A**	A	C	**C**	C	G	C	T	A	G	A	**C**	G
Hap70	*41		**C**	**C**	**G**	G	C	**C**	C	A	T	T	A	G	A	**C**	G
Hap71	*41		**C**	C	G	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap72	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap73	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap74	*41		**C**	C	G	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap75	*41		**C**	C	G	**G**	C	**C**	C	**A**	**T**	T	A	G	A	**C**	**A**
Hap76	*41		**C**	C	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap77	*41		**C**	C	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap78	*41		**C**	C	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap79	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap80	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap81	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	**A**
Hap82	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap83	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	G
Hap84	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	**C**	A	G	A	**C**	**A**
Hap85	*41		**C**	C	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	**A**
Hap86	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap87	*41		**C**	**T**	**A**	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	G
Hap88	*41		**C**	C	G	A	C	**C**	C	**A**	**T**	T	A	G	A	**C**	**A**
Hap89	*65		**C**	C	G	**G**	C	**C**	C	G	**T**	T	**G**	G	A	**C**	G
Hap90	ndt		**C**	C	G	**G**	C	**C**	C	G	**T**	T	A	G	A	**C**	G
Hap91	ndt		G	C	G	A	C	A	C	**A**	**T**	T	A	G	A	G	G
Hap92	ndt		**C**	**T**	**A**	A	C	**C**	C	**A**	C	T	A	G	A	**C**	G
Hap93	ndt		**C**	**T**	**A**	A	C	**C**	C	**A**	C	T	A	G	A	**C**	G

rs368389952	rs1081003	rs28371705	rs28371704	rs29001678	rs1081000	rs1065852	rs138100349	rs769258	rs372204775	rs530422334	rs374672076	rs566383351				
6089G>A	6057C>T	6015C>G	6002A>G	5289C>T	5264A>G	5119C>T	5101C>T	5050G>A	4818G>A	4666A>G	4622G>C	4589C>T	Balkan Roma	Baranja Roma	Medjimurje Roma	Combined
G	C	C	A	C	A	C	C	G	G	A	G	**T**	3	0	0	3
A	C	C	A	C	A	C	C	G	G	A	G	C	9	0	0	9
A	C	C	A	C	A	C	C	G	G	A	G	**T**	1	0	0	1
G	C	C	A	C	A	C	C	G	G	A	G	**T**	0	1	0	1
G	C	C	A	C	A	C	C	G	G	**G**	G	C	0	1	0	1
G	C	C	A	**T**	A	C	C	G	G	A	G	C	0	3	0	3
G	C	C	A	C	A	C	C	G	G	A	G	**T**	0	1	0	1
G	C	C	A	C	A	C	C	G	G	A	G	C	0	1	0	1
G	C	C	A	C	A	C	C	G	G	A	G	C	26	55	43	124
G	C	C	A	C	A	C	C	G	G	A	**C**	C	0	0	1	1
G	C	C	A	C	A	C	C	G	G	A	G	C	1	6	3	10
G	C	C	A	C	A	C	C	G	G	A	G	C	4	14	28	46
G	C	C	A	C	A	C	C	G	G	A	G	**T**	6	1	2	9
G	C	C	A	**T**	A	C	C	G	G	A	G	C	0	1	1	2
G	C	C	A	C	A	C	C	G	G	A	G	C	1	0	0	1
G	C	C	A	C	A	C	C	G	G	A	G	C	2	0	0	2
G	C	C	A	C	A	C	C	G	G	A	G	**T**	5	0	0	5
G	C	C	A	C	A	C	C	G	G	A	G	C	1	0	0	1
A	C	C	A	C	A	C	C	G	G	A	G	C	1	0	0	1
A	C	C	A	C	A	C	C	**A**	G	A	G	C	1	0	0	1
G	C	C	A	**T**	A	C	C	G	G	A	G	C	0	1	0	1
G	C	C	A	C	A	C	C	G	G	A	G	**T**	0	1	0	1
G	C	C	A	C	A	C	C	G	**A**	A	G	C	0	1	0	1
G	C	C	A	C	**G**	C	C	G	G	A	G	**T**	5	1	0	6
G	C	C	A	C	A	C	C	G	G	A	G	C	6	3	10	19
G	C	C	A	C	A	C	C	G	G	A	G	**T**	0	0	2	2
G	C	C	A	C	A	C	C	G	G	A	G	C	21	21	26	68
G	C	C	A	C	A	C	C	G	G	A	G	**T**	6	14	17	37
G	C	C	A	C	A	C	C	G	G	A	G	C	0	0	1	1

G	C	C	A	C	A	C	C	G	G	**G**	G	**T**	0	0	1	1
G	C	C	A	C	A	C	C	G	G	A	G	C	0	0	1	1
A	C	C	A	C	A	**T**	C	G	G	A	G	C	3	0	0	3
A	C	C	A	C	A	**T**	C	G	G	A	**C**	C	2	0	0	2
G	C	C	A	C	A	**T**	C	G	G	A	G	C	1	0	0	1
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	5	0	0	5
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	1	0	0	1
G	C	G	**G**	C	A	**T**	C	G	G	A	G	C	7	0	0	7
G	C	**G**	**G**	C	A	**T**	C	G	G	A	**C**	C	0	2	0	2
G	C	C	A	C	A	**T**	C	G	G	A	G	C	0	2	0	2
G	C	C	A	C	A	**T**	C	G	G	A	G	C	0	1	0	1
G	C	**G**	**G**	C	A	**T**	C	G	**A**	A	**C**	C	0	7	0	7
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	2	3	0	5
G	C	C	A	C	A	**T**	C	G	G	A	G	C	1	1	0	2
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	5	8	0	13
G	C	C	A	C	A	**T**	C	G	G	**G**	**C**	C	0	3	1	4
G	C	C	A	C	A	**T**	C	G	**A**	A	G	C	0	4	7	11
G	C	C	A	C	A	**T**	C	G	**A**	A	**C**	C	2	0	2	4
G	C	C	A	C	A	C	C	G	G	A	**C**	C	0	0	2	2
G	C	C	A	C	A	**T**	C	G	G	A	G	C	0	0	1	1
G	C	C	A	C	A	**T**	C	G	G	A	G	C	0	0	1	1
G	C	C	A	C	A	C	C	G	G	A	**C**	C	0	0	3	3
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	0	0	1	1
G	C	C	A	C	A	**T**	C	G	G	A	G	C	19	17	16	52
G	C	C	A	C	A	**T**	C	G	G	A	G	C	0	1	0	1
G	**T**	C	A	C	A	**T**	C	G	G	A	G	C	0	1	0	1
G	C	C	A	C	A	**T**	C	G	G	A	**C**	C	0	1	0	1
G	C	C	A	C	A	**T**	C	G	**A**	A	G	C	0	1	0	1
G	C	C	A	C	A	**T**	C	G	G	A	G	C	2	6	9	17
G	**T**	C	A	C	A	**T**	C	G	G	A	G	C	4	2	8	14
G	**T**	C	A	C	A	**T**	C	G	G	**G**	G	C	0	0	4	4
G	C	C	A	C	A	C	**T**	G	G	A	G	C	1	0	2	3
G	C	C	A	**T**	A	C	C	G	G	A	G	C	2	0	0	2

G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	0	2	2
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	1	1	2
G	C	C	A	C	A	C	**A**	G	A	G	C	C	C	**A**	A	C	C	C	2	2	0	4
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	1	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	0	1	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	1	1	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	3	3	0	6
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	1	0	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	2	0	0	2
G	C	C	A	C	A	C	G	G	**G**	G	C	C	C	G	A	C	C	C	4	0	0	4
A	C	C	A	C	A	C	G	G	A	G	C	C	C	**T**	A	C	C	C	1	0	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	1	0	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	1	0	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	2	6	0	8
G	C	C	A	C	A	C	G	G	A	G	C	C	C	**T**	A	C	C	C	0	2	0	2
G	C	C	A	C	A	**T**	G	G	A	G	C	C	C	**T**	A	C	C	C	0	1	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	**T**	A	C	C	C	5	3	0	8
G	C	C	A	C	A	C	G	G	A	G	C	C	C	**T**	A	C	C	C	0	6	0	6
G	C	C	A	C	A	C	G	G	A	G	C	**T**	C	G	A	C	C	C	0	1	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	0	1	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	13	18	10	41
G	C	C	A	C	A	**T**	G	G	A	G	C	C	C	G	A	C	C	C	0	0	1	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	0	1	1
G	C	C	A	C	A	C	G	G	A	G	**T**	C	C	G	A	C	C	C	0	1	0	1
G	C	C	A	C	A	C	G	G	A	G	**T**	C	C	G	A	C	C	C	1	0	0	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	**T**	2	0	0	2
G	C	C	A	C	A	C	G	G	A	G	C	C	C	G	A	C	C	C	0	0	1	1
G	C	C	A	C	A	C	G	G	A	G	C	C	C	**G**	**G**	C	C	**T**	0	0	1	1

Supp. Table 3 - Worldwide frequencies of most frequent star allels in Roma population

	*1			*2			*4			*10			*39			*41		
Population	average	min	max	average	min	max	average	min	max	average	min	max	average	min	max	average	min	max
African Americans*	33,94240653	30,6	33,96777778	12,84516061	4,2	28,7	5,933902098	3,86	8	3,834007372	2,7	7,5	n/a	n/a	n/a	11,20730491	1,84	14,9
Africa*	32,72393976	6,75	56,1	20,43211446	10,6	40	3,664840668	0,9	7,07	6,284542203	2,53	19,23	0,688468158	0	1,6	9,012431562	1,44	25,3
Americas*	48,02099868	10,29	73,5	23,54009588	6,1	41	10,90404322	0,2	24	3,155263585	0	12,45	0,723404255	0	0,8	4,309691664	0	13,47
East Asia*	36,85679852	17,5	93,79	11,8773296	7,65	42,71	0,583108981	0	4,35	42,17881563	8,6	64,1	0,620156352	0	1,18	2,431460237	0	6,54
Europe*	38,08394218	18,5	65,79	24,9025069	10,53	40,63	18,8064618	8,1	33,4	2,196179205	0,38	8	4,560970232	0	14,6	8,512889495	2,7	12,4
Middle East*	35,23513057	27,65	49	19,33070661	9	32	8,131103286	3,5	12,5	3,395132979	0,7	9	2,684444444	0	4	17,0832	15,2	29
Oceania*	76,01333333	60	85,4	2,204888889	0	3,8	1,356666667	0	3	1,210222222	0	3	0	0	0	0	0	0
South/Central Asia*	50,90245595	39,6	84,2	31,55543547	12,5	38,3	9,078790297	2,8	18,44	13,59213483	3,8	57,7	0,2			9,964454976	0	12,5
Croatian Roma	33,8	26,6	37,3	23,76	18,26	27,75	20,73	16,27	25,53	6,22	3,91	10,04	1,554404145			14,04	6,22	18,45
Balkan	25,90673575			25,38860104			24,87046632			3,10880829			1,554404145			16,58031088		
Baranja	35,8974359			17,94871795			20,51282051			5,128205128			0,427350427			18,37606838		
Medjimurje	36,44859813			27,10280374			15,88785047			9,813084112			2,336448598			6,074766355		

*Data from Geadigk et al. 2017 with pondered average values

2.2. FROM CROATIAN ROMA TO 1000 GENOMES: THE STORY OF THE *CYP2D6* GENE PROMOTER AND ENHANCER SNPs

Journal of *Personalized Medicine*

MDPI

Article

From Croatian Roma to 1000 Genomes: The Story of the *CYP2D6* Gene Promoter and Enhancer SNPs

Anita Stojanović Marković, Željka Celinšćak, Maja Šetinc, Tatjana Škarić-Jurić, Marijana Peričić Salihović and Matea Zajc Petranović *

Institute for Anthropological Research, 10000 Zagreb, Croatia
* Correspondence: matea@inantro.hr

Abstract: The *CYP2D6* gene encodes an enzyme responsible for the metabolism of ~20% of clinically prescribed drugs. In this study, 18 SNPs from the enhancer and promoter regions of *CYP2D6* in 323 Roma from Croatia were genotyped, to find out whether the demographic history of Roma affected the distribution of the studied SNPs and their linkage disequilibrium (LD) values, with the major SNPs defining the *CYP2D6* star alleles. No differences were found between the three Roma groups in allele and genotype frequencies. The distribution of LD values of Roma was compared with LD values of European and Asian populations. Regulatory *CYP2D6* SNPs (rs5758550, rs28624811, rs1080985 and rs1080983) showed similar distribution and the highest LDs with rs16947 from the gene-coding region in all populations. In the promoter region, a complete LD between rs1080989 and rs28588594, and between rs1080983 and rs28624811, was found in Croatian Roma and investigated populations from 1000 genomes. A high LD was also found between rs1080985 from the promoter and rs5758550 from the enhancer region. SNP rs28735595 from the gene promoter region had the highest LD, with two gene region SNPs, rs1058164 and rs1135840. To conclude, the Croatian Roma population shows an LD pattern of the *CYP2D6* gene region similar to the 1000 Genomes European and Asian populations.

Keywords: *CYP2D6* gene; population genetics; Roma population; promoter; enhancer; regulation of transcription; pharmacogenomics; personalized medicine

Citation: Stojanović Marković, A.; Celinšćak, Ž.; Šetinc, M.; Škarić-Jurić, T.; Peričić Salihović, M.; Zajc Petranović, M. From Croatian Roma to 1000 Genomes: The Story of the *CYP2D6* Gene Promoter and Enhancer SNPs. *J. Pers. Med.* **2022**, *12*, 1353. https://doi.org/10.3390/jpm12081353

Academic Editor: George P. Patrinos

Received: 19 July 2022
Accepted: 17 August 2022
Published: 22 August 2022

1. Introduction

The *CYP2D6* gene encodes a homonymous drug-metabolizing cytochrome P450 enzyme, responsible for eliminating more than 21% of clinically used drugs [1]. This gene, located on chromosome 22q13.1, is highly polymorphic and its genetic variations greatly contribute to the inter-individual variability of CYP2D6 enzyme activity, which is divided into four categories: poor metabolizer, intermediate metabolizer, normal metabolizer and ultrarapid metabolizer [2–5]. The Clinical Pharmacogenetics Implementation Consortium (CPIC) offers guidelines for assigning activity scores of *CYP2D6* variant alleles, and subsequent translations of diplotypes into phenotypes—this method was proposed and established by Gaedigk and colleagues to standardize genotype-to-phenotype translations [6]. However, recent studies have shown that even in individuals with the same genotype, CYP2D6 enzyme activity can vary up to several times [7–10], and differential regulation of *CYP2D6* transcription may partly explain the variability in *CYP2D6*-mediated drug metabolism [11].

According to the GeneCards database [12], there are 87 loci in enhancer and promoter regions related to the expression of the *CYP2D6* gene, spanning from less than 1000 to almost 300,000 base pairs away from the transcription starting site (TSS). Among these, eight are active in hepatocytes (https://epd.epfl.ch/, https://www.encodeproject.org/, and https://www.ncbi.nlm.nih.gov/refseq/ (accessed on 15 March 2022)). The association of enhancer/promoter activity and variations in the *CYP2D6* gene with overall drug metabolizing is not extensively studied. So far, only a few regulatory variants of the *CYP2D6* gene

have been studied. Mostly, the two completely linked single-nucleotide polymorphisms (SNPs), rs133333 (G > A) and rs5758550 (G > A), located ~116 kb downstream of the gene and identified as enhancers [13]. These SNPs are located within the 2.4 kb-long enhancer GH22J042015, the binding site for the transcription factor ZNF512 [12]. In addition to the aforementioned enhancer SNPs, the other most studied SNPs are located within the promoter/enhancer GH22J042130, which is 1.6 kb in size and 0.7 kb away from the TSS of the *CYP2D6* gene. According to the Pharmacogene Variation (PharmVar) Consortium, some of the SNPs from this region are part of haplotypes that define the *CYP2D6* star alleles [14–16].

Considering all cell types, GH22J042130 is the biding site of 15 transcription factors and affects the transcription of 11 genes [12]. In transcription related to drug metabolizing activity, this promoter is induced by the binding of hepatocyte nuclear factor 4 alpha, Kruppel-like factor 9 and peroxisome proliferator-activated receptor alpha, and is suppressed by nitric oxide and estrogen [11]. Since studies have shown that haplotypes containing enhancer or promoter loci allow the determination of CYP2D6 enzyme activity in vivo, their inclusion in genotyping panels could allow more accurate prediction of CYP2D6 activity [12]. SNPs from enhancers and promoter regions may be in linkage disequilibrium (LD) with star allele-defining SNPs from the *CYP2D6* gene region, which may influence the metabolizing effect [17]. LD differs among populations, especially if they are isolated, have a different ancestry from the surrounding majority population and are susceptible to genetic drift.

An example of such a population is the Roma (Gypsy) population, a transnational minority present in many countries around the world. They originated in India and arrived in Europe around the 11th century via Central Asia (Afghanistan and Persia), the Middle East and present-day Turkey. It is estimated that the Roma population is numbering around 15 million people worldwide, of whom 12 million reside in Europe. Roma in Croatia belong to two socio-culturally and linguistically different groups: Vlax Roma, descendants of the Roma who crossed the Danube River between the 13th and 15th century and arrived in Wallachia and Transylvania (both in present-day Romania), and Moldavia, where they were forced to work in the mines for the next 500 years. During that time, they were forbidden to use their own language, so their descendants are now recognized by a specific archaic Romanian language—ljimb'd bayash. The second group is Balkan Roma, descendants of the Roma who arrived in the Balkans in the 11th century, and they speak dialects of the romani chib language. Socio-cultural characteristics of the Roma population, such as strict rules of endogamy, in addition to the founder and the bottleneck effects, have caused the genetic structure of Roma to differ compared to other populations [18–20], which has been shown to affect ADME genes' variations as well [21].

The main objective of this study was to estimate the variation in enhancers and promoter regions of the *CYP2D6* gene among: (a) three socio-culturally and geographically distinct Croatian Roma groups, and (b) Croatian Roma and European and Asian 1000 Genomes populations; in particular, to find out whether the specific history of the Roma population influenced the distribution of the studied SNPs and their LDs with the main *CYP2D6* star allele-defining SNPs. The knowledge of LD between *CYP2D6* star allele-defining SNPs and SNPs in promoter regions and/or enhancers can enable prediction of *CYP2D6* activity with greater accuracy.

2. Materials and Methods

We analyzed 323 DNA samples, all collected during field studies of the ongoing multidisciplinary anthropological, molecular-genetic and epidemiological investigations of Roma populations in Croatia. Samples belong to members of the three socio-culturally different Roma subpopulations: the Vlax Roma, who are divided into two subpopulations according to the geographical regions of Croatia they inhabit: Baranja and Medjimurje, and the Balkan Roma from the city of Zagreb. All Roma participated in the study voluntarily, and with the help of Roma volunteers, were informed about the goals, methods and

expectations of the study. The Scientific Board and the Ethics Committee of the Institute for Anthropological Research in Zagreb, Croatia, approved the study protocol.

DNA was extracted from peripheral blood using the salting-out method [22]. The genotyping of 16 SNPs in the promoter region of the *CYP2D6* gene and two enhancer SNPs on chromosome 22 was carried out using the Kompetitive Allele-Specific PCR method (KASP) in a commercial facility. The KASP genotyping assay is a form of competitive allele-specific PCR combined with a homogeneous fluorescent SNP genotyping system, which determines the alleles at a specific locus within genomic DNA [23]. Data for the *CYP2D6* star allele-defining SNPs (rs1135840, rs16947, rs28371725, rs3892097, rs1058164, rs1065852 and rs769258) in Croatian Roma were taken from a paper by Stojanović Marković et al. [24].

Allele and genotype frequencies were calculated by direct counting. Hardy–Weinberg equilibrium (HWE) was assessed using the software Arlequin 3.5 [25]. Genotype and allele frequency differences between the three Roma groups were tested using the Chi-square test. The analyses were performed using R with statistical significance set at $p < 0.05$ [26]. Linkage disequilibrium (LD) analyses in the Roma groups have been performed using the software Haploview [27]. Only r^2 values of LD were calculated since it is considered more robust than D′ and correlates better among different population samples [28,29]. Haploview software was also used for drawing plots. Data from the 1000 Genomes database were used to compare the Croatian Roma population with European and Asian populations for the SNPs studied. The European cluster consisted of the following populations: Utah residents with Northern and Western ancestry (CEU), Finland (FIN), British in England and Scotland (GBR), Iberian population in Spain (IBS) and Toscani in Italy (TSI). The East Asian cluster consisted of Dai Chinese (CDX), Han Chinese in Beijing (CHB), South Han Chinese (CHS), Japanese in Tokyo (JPT) and Kinh in Ho Chi Minh City, Vietnam (KHV), while the South Asian cluster consisted of Bengali in Bangladesh (BEB), Gujarati Indian (GIH), Indian Telugu in the UK (ITU), Punjabi in Lahore Pakistan (PJL) and Sri Lankan Tamil in the UK (STU). LDs for European, East Asian and South Asian populations were calculated using the LD calculator implemented in the Ensembl genome browser [30]. Spearman's correlation was used to compare LD values [26]. Spearman's correlation results were used as input for multidimensional scaling (MDS), and plots were drawn using ggplot2 [31].

3. Results

Allele and genotype frequencies of studied polymorphic sites determined in three Croatian Roma subpopulations are shown in Table 1. Eight out of the eighteen investigated SNPs in our sample were monomorphic (rs1080993, rs34894147, rs1376235338, rs1224722684, rs1409156443, rs536645539, rs1080990 and rs58188898). All polymorphic sites except for rs13333 in the Baranja Roma subpopulation were in Hardy–Weinberg equilibrium. None of the SNPs showed significant differences in genotype or allele frequencies between the three Roma groups (Table 1).

Table 1. Genotype and allele frequencies of 16 *CYP2D6* promoter and 2 enhancer SNPs in the three Croatian Roma samples (Baranja, Medjimurje and Balkan).

Polymorphisms		Genotypes and Alleles	Baranja	Medjimurje	Balkan	Total	Chi Square	*p*	HWE Baranja	HWE Medjimurje	HWE Balkan
rs133333	genotypes	A/A	78	53	56	187	8.298	0.081	0.038	0.739	0.303
		A/G	24	39	30	93					
		G/G	6	6	7	19					
	alleles	A	180	145	142	467	5.738	0.057			
		G	36	51	44	131					
rs5758550	genotypes	A/A	78	54	56	188	8.186	0.085	0.092	0.499	0.491
		A/G	27	43	33	103					
		G/G	6	6	7	19					
	alleles	A	183	151	145	479	5.549	0.062			
		G	39	55	47	141					

Table 1. *Cont.*

Polymorphisms		Genotypes and Alleles	Baranja	Medjimurje	Balkan	Total	Chi Square	*p*	HWE Baranja	HWE Medjimurje	HWE Balkan
rs1080993	genotypes	C/C	113	107	93	313					
	alleles	C	226	214	186	626					
rs34894147	genotypes	CC/CC	112	104	96	312					
	alleles	CC	224	208	192	624					
rs1376235338	genotypes	C/C	114	104	96	314					
	alleles	C	228	208	192	628					
rs35046171	genotypes	G/G A/G	113 1	104 0	95 1	312 2	1.019	0.601	0.963		0.959
	alleles	G A	227 1	208 0	191 1	626 2	1.016	0.602			
rs1224722684	genotypes	G/G	114	105	96	315					
	alleles	G	228	210	192	630					
rs34167214	genotypes	A/A C/A	113 1	105 1	94 0	312 2	0.864	0.649	0.963	0.961	
	alleles	A C	227 1	211 1	188 0	626 2	0.861	0.650			
rs1409156443	genotypes	C/C	113	106	96	315					
	alleles	C	226	212	192	630					
rs28624811	genotypes	G/G G/A A/A	45 44 20	40 51 13	34 41 20	119 136 53	3.573	0.467	0.123	0.598	0.251
	alleles	G A	134 84	131 77	109 81	374 242	1.392	0.499			
rs536645539	genotypes	TC/TC	112	106	93	311					
	alleles	TC	224	212	186	622					
rs1080990	genotypes	C/C	114	104	94	312					
	alleles	C	228	208	188	624					
rs1080989	genotypes	C/C C/T T/T	60 45 8	58 35 10	47 35 9	165 115 27	1.335	0.855	0.912	0.181	0.515
	alleles	C T	165 61	151 55	129 53	445 169	0.335	0.846			
rs28735595	genotypes	C/C C/T T/T	46 51 14	39 45 17	45 39 8	130 135 39	3.900	0.420	0.981	0.517	0.913
	alleles	C T	143 79	123 79	129 55	395 213	3.642	0.162			
rs28588594	genotypes	G/G G/A A/A	60 45 9	60 35 10	49 35 9	169 115 28	1.074	0.898	0.890	0.158	0.461
	alleles	G A	165 63	155 55	133 53	453 171	0.273	0.873			
rs1080985	genotypes	G/G C/G C/C	75 29 7	55 43 6	53 32 6	183 104 19	5.757	0.218	0.085	0.523	0.697
	alleles	G C	179 43	153 55	138 44	470 142	3.153	0.207			

Table 1. *Cont.*

Polymorphisms		Genotypes and Alleles	Baranja	Medjimurje	Balkan	Total	Chi Square	*p*	HWE Baranja	HWE Medjimurje	HWE Balkan
rs58188898	genotypes	G/G	111	106	96	313					
	alleles	G	222	212	192	626					
rs1080983	genotypes	C/C T/C T/T	49 45 20	42 50 12	35 41 19	126 136 51	3.769	0.438	0.096	0.619	0.277
	alleles	C T	143 85	134 74	111 79	388 238	1.601	0.449			

The two *CYP2D6* gene enhancer SNPs are highlighted in grey, while the other SNPs are from the promoter region. Significant Chi-square and HWE *p*-values are shown in bold.

In Table 2, the linkage disequilibrium (LD) values for the three Croatian Roma subpopulations (r^2 values) are shown for pairs of two enhancer (rs133333 and rs5758550) and six polymorphic promoter SNPs (rs28624811, rs1080989, rs28735595, rs28588594, rs1080985 and rs1080983), as well as between pairs of the latter and SNPs that define different *CYP2D6* gene star alleles (rs1135840, rs16947, rs28371725, rs3892097, rs1058164, rs1065852 and rs769258). Two SNPs in the *CYP2D6* gene promoter region, rs35046171 and rs34167214, were not included in the LD calculation due to the extremely low prevalence of minor alleles in these SNPs.

Table 2. LD values (r^2) between pairs of polymorphic sites in the *CYP2D6* gene regulatory and gene-coding regions in the three Croatian Roma groups (Baranja, Medjimurje and Balkan).

		Baranja	Medjimurje	Balkan
L1	**L2**	**r^2**	**r^2**	**r^2**
rs133333	rs5758550	**1**	**1**	**1**
rs133333	rs1135840	0.116	0.210	0.094
rs133333	rs28371725	0.046	0.030	0.051
rs133333	rs16947	0.350	0.527	0.322
rs133333	rs3892097	0.056	0.071	0.104
rs133333	rs1058164	0.113	0.175	0.098
rs133333	rs1065852	0.075	0.114	0.089
rs133333	rs769258	0.047		0.039
rs133333	rs28624811	0.344	0.618	0.437
rs133333	rs1080989	0.077	0.131	0.119
rs133333	rs28735595	0.111	0.218	0.123
rs133333	rs28588594	0.079	0.131	0.123
rs133333	rs1080985	**0.877**	**0.920**	**0.937**
rs133333	rs1080983	0.347	0.642	0.430
rs5758550	rs1135840	0.121	0.224	0.100
rs5758550	rs28371725	0.050	0.029	0.052
rs5758550	rs16947	0.353	0.551	0.333
rs5758550	rs3892097	0.057	0.072	0.105
rs5758550	rs1058164	0.118	0.188	0.103
rs5758550	rs1065852	0.077	0.114	0.092
rs5758550	rs769258	0.043		0.036

Table 2. *Cont.*

		Baranja	Medjimurje	Balkan
L1	**L2**	**r^2**	**r^2**	**r^2**
rs5758550	rs28624811	0.353	0.635	0.442
rs5758550	rs1080989	0.079	0.130	0.119
rs5758550	rs28735595	0.115	0.232	0.130
rs5758550	rs28588594	0.081	0.130	0.123
rs5758550	rs1080985	**0.884**	**0.925**	**0.941**
rs5758550	rs1080983	0.350	0.659	0.436
rs1135840	rs28624811	0.312	0.342	0.241
rs1135840	rs1080989	0.147	0.197	0.178
rs1135840	rs28735595	**0.922**	**0.958**	**0.890**
rs1135840	rs28588594	0.153	0.194	0.171
rs1135840	rs1080985	0.114	0.227	0.086
rs1135840	rs1080983	0.315	0.345	0.262
rs28371725	rs28624811	0.345	0.153	0.251
rs28371725	rs1080989	0.082	0.031	0.089
rs28371725	rs28735595	0.132	0.056	0.043
rs28371725	rs28588594	0.084	0.030	0.090
rs28371725	rs1080985	0.054	0.028	0.056
rs28371725	rs1080983	0.379	0.146	0.252
rs16947	rs28624811	**0.921**	**0.807**	**0.861**
rs16947	rs1080989	0.210	0.183	0.296
rs16947	rs28735595	0.303	0.289	0.175
rs16947	rs28588594	0.214	0.178	0.303
rs16947	rs1080985	0.346	0.491	0.283
rs16947	rs1080983	**0.943**	0.787	**0.860**
rs3892097	rs28624811	0.161	0.115	0.249
rs3892097	rs1080989	0.710	0.543	**0.842**
rs3892097	rs28735595	0.114	0.116	0.141
rs3892097	rs28588594	0.698	0.544	**0.844**
rs3892097	rs1080985	0.043	0.070	0.107
rs3892097	rs1080983	0.159	0.108	0.245
rs1058164	rs28624811	0.303	0.312	0.248
rs1058164	rs1080989	0.145	0.222	0.183
rs1058164	rs28735595	**0.903**	**0.959**	**0.918**
rs1058164	rs28588594	0.151	0.218	0.176
rs1058164	rs1080985	0.111	0.192	0.090
rs1058164	rs1080983	0.306	0.312	0.270
rs1065852	rs28624811	0.219	0.186	0.269
rs1065852	rs1080989	**0.956**	**0.880**	**0.973**

Table 2. *Cont.*

		Baranja	Medjimurje	Balkan
L1	**L2**	**r^2**	**r^2**	**r^2**
rs1065852	rs1080985	0.066	0.070	0.096
rs1065852	rs1080983	0.217	0.175	0.264
rs769258	rs28624811	0.015		0.015
rs769258	rs1080989	0.003		0.002
rs769258	rs28735595	0.005		0.001
rs769258	rs28588594	0.003		0.001
rs769258	rs1080985	0.038		0.039
rs769258	rs1080983	0.015		0.016
rs28624811	rs1080989	0.226	0.210	0.284
rs28624811	rs28735595	0.332	0.360	0.309
rs28624811	rs28588594	0.230	0.213	0.290
rs28624811	rs1080985	0.371	0.567	0.401
rs28624811	rs1080983	**0.981**	**1**	**1**
rs1080989	rs28735595	0.167	0.190	0.181
rs1080989	rs28588594	**1**	**1**	**1**
rs1080989	rs1080985	0.068	0.095	0.127
rs1080989	rs1080983	0.223	0.206	0.282
rs28735595	rs28588594	0.172	0.190	0.170
rs28735595	rs1080985	0.134	0.231	0.096
rs28735595	rs1080983	0.336	0.360	0.307
rs28588594	rs1080985	0.071	0.091	0.127
rs28588594	rs1080983	0.227	0.200	0.285
rs1080985	rs1080983	0.376	0.582	0.404

High LD values ($r^2 > 0.8$) are shown in bold.

Figure 1 graphically shows the distribution of LD values between the *CYP2D6* promoter region/enhancers' SNPs and star allele-defining SNPs in the European, South Asian, East Asian and Croatian Roma populations.

In the studied world populations, four SNPs in the *CYP2D6* regulatory regions (rs5758550 in the enhancer region, and rs28624811, rs1080985 and rs1080983 in the promoter region) showed similar distributions and the highest LD with rs16947 from the *CYP2D6* gene region (Figure 1). Since these four regulatory region SNPs have the same LD pattern with the SNPs in the gene region, we calculated their pairwise LDs and found that promoter regions rs1080983 and rs28624811 are in complete LD not only in the European and Asian populations (data not shown), but also in the Croatian Roma population (Table 2). A high LD ($r^2 > 0.8$) was found between rs1080985, from the promoter region, and rs5758550, from the enhancer region, both in world populations and Croatian Roma groups (Figure 1). Other SNP pairs have LD values ranging from 0.4 to 0.8, with the highest values in the Finnish population. Promoter region SNPs rs1080989 and rs28588594 also showed nearly identical distribution of LD values in the studied populations, and the highest LDs with rs1065852 from the *CYP2D6* gene-coding region. We tested LD values between rs1080989 and rs28588594 from the promoter region and found that they were in complete LD in all studied populations. SNP rs28735595, from the *CYP2D6* gene promoter region, had the highest LD with two SNPs in the gene region, rs1058164 and rs1135840 ($r^2 > 0.8$ for both SNPs). A more precise insight into LD values in Croatian Roma groups is shown in the Supplementary Figures S1–S3, which show nine LD plots for each of the Croatian Roma subpopulations.

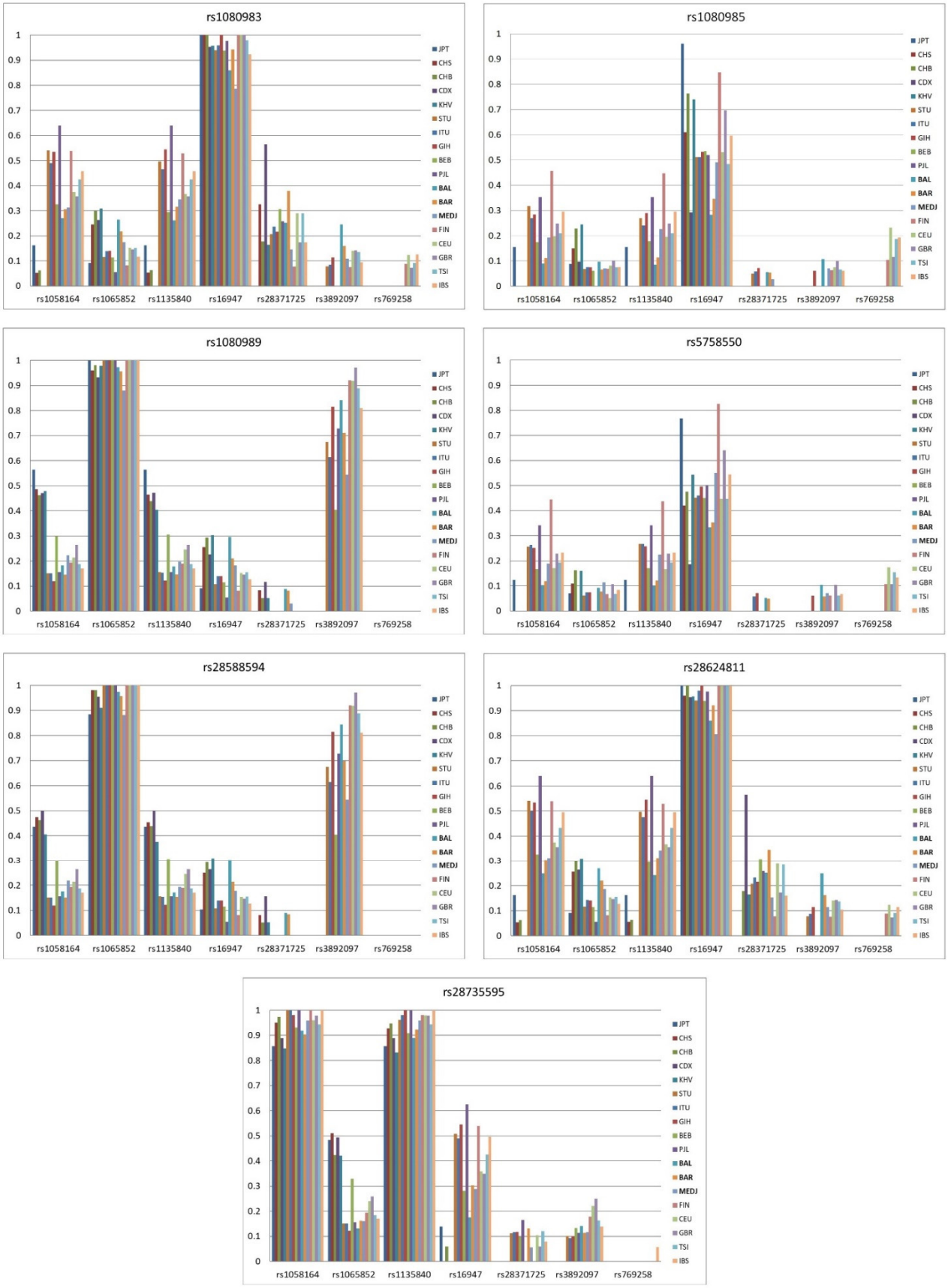


Figure 1. LD values' (r^2) distribution between the *CYP2D6* promoter and enhancer SNPs and star allele-defining SNPs in the European, South Asian, East Asian and Croatian Roma populations.

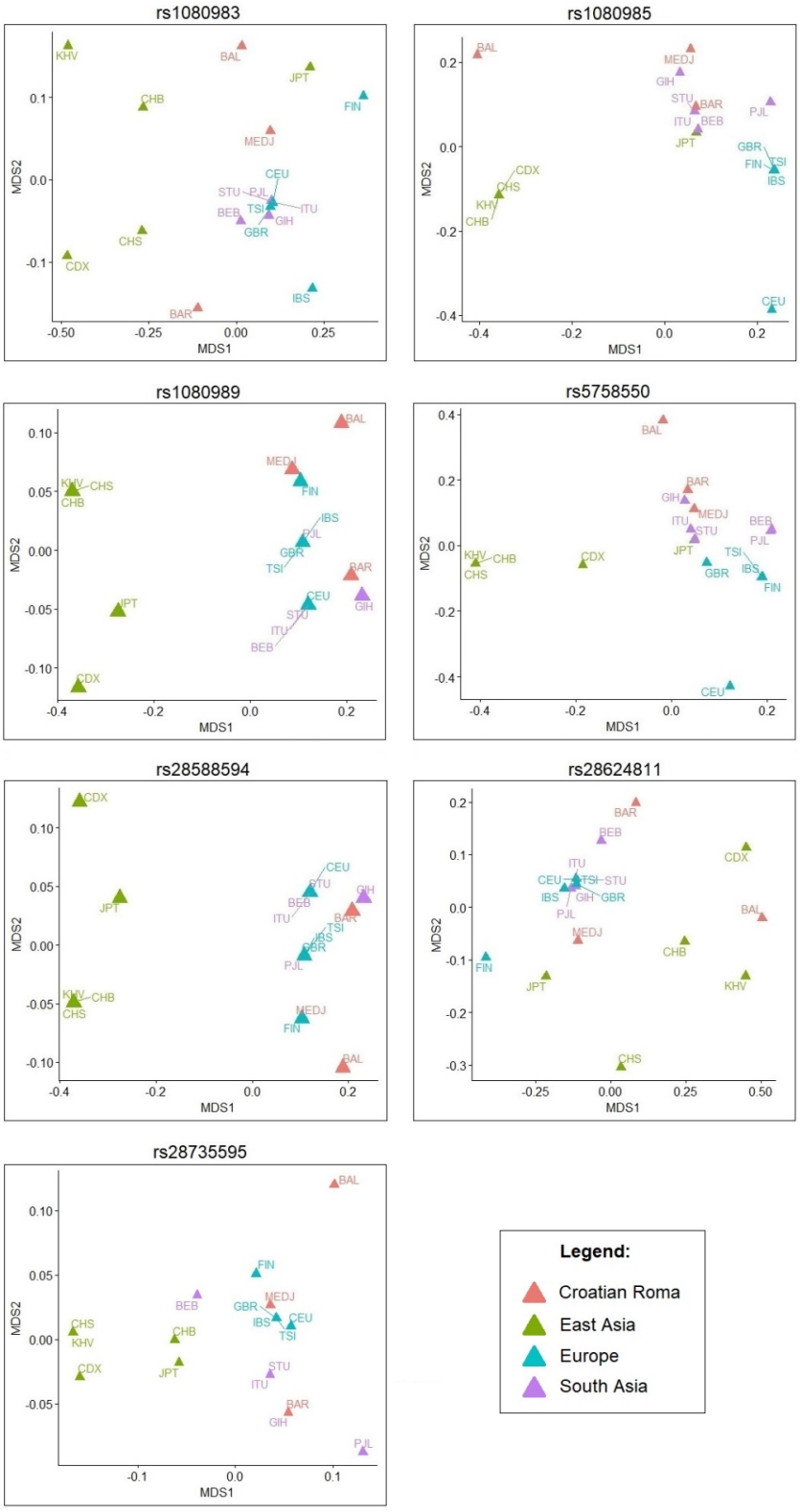


Figure 2. Multidimensional scaling plots (MDS) of Spearman's correlation matrices for linkage disequilibrium (LD) between the *CYP2D6* promoter and enhancer SNPs and star allele-defining SNPs in the European, South Asian, East Asian and Croatian Roma populations.

To reveal the pattern of LD correlations, the MDS plots (Figure 2) were constructed as described in the Materials and Methods Section. Most MDS plots show separation of East Asian populations from other populations. Considering the Croatian Roma population, the

plots also suggest a slightly remote position of the Balkan Roma subpopulation from others, while the Baranja Roma subpopulation is almost always positioned close to some of the South Asian populations. The Roma subpopulation from Medjimurje is positioned either close to South Asian populations (MDS plots for rs5758550, rs28624811 and rs1080985), or closer to European populations (MDS plots for rs28735595, rs28588594 and rs1080989).

4. Discussion

Population pharmacogenomics is a growing area driven by increasing population data on genes responsible for absorption, distribution, metabolism and excretion (ADME genes). Population ancestry may affect the diversity of genetic polymorphisms, leading to population-specific differences in drug responses [32]. Within population pharmacogenomics, special attention should be given to the study of indigenous and/or minority populations which, due to their genetic history, show a specific distribution of alleles that can alter drug metabolism and lead to adverse drug reactions (ADR).

The pharmacogenomics of the Roma minority population has been studied for the last few years [33]. These studies included SNPs in several ADME genes, such as *ABCB1* [34,35], *CYP2B6* [36–38], *CYP2C19* [39–42], *CYP2D6* [24,37,38,41] and *NAT* [42,43].

Previous analyses showed that the three socio-culturally different Croatian Roma groups show significant differences in allele distribution within the *CYP2D6* gene [24], and therefore we continued to investigate promoter and enhancer SNPs associated with this gene. In general, diversity in regulatory elements has an impact on gene expression, so understanding it could help to elucidate the unexplained variability in gene activity [11]. Contrary to the differences found among Croatian Roma groups in the *CYP2D6* gene region, the regulatory elements studied here showed no difference among the same Roma subpopulations.

To clarify the relationships of SNPs in the promoter/enhancer region with star allele-defining SNPs from the *CYP2D6* gene region in the Croatian Roma population, we determined their LDs. Significant LDs between SNPs in regulatory and gene regions may affect *CYP2D6* transcription and consequently drug metabolism, and so far, the most studied example of this interaction is rs5758550 [44]. Using the reporter gene assay, Wang et al. [45] found that the constructs containing minor allele G had higher activity independently of other SNPs which were part of the construct (rs133333 and rs4822082), while deletion of the region surrounding rs5758550 decreased *CYP2D6* mRNA levels. Rs133333 and rs5758550 are in complete LD, but chromatin immunoprecipitation with the P300 antibody showed that deletion of 156 bp surrounding rs133333 did not decrease the level of transcription of *CYP2D6* [45]. SNPs rs5758550 and rs133333, genotyped in Croatian Roma subpopulations, were also in high LD. The LDs of enhancers rs5758550 and rs133333 with SNPs from the *CYP2D6* promoter region were also calculated. Only rs1080985 was in LD with the two enhancer SNPs. Raimundo et al. [46] and Zanger et al. [47] linked rs1080985 with increased *CYP2D6* expression in the human liver, but this was not supported by reporter gene assays [13]. Today, it is considered that this SNP has no functional consequences (https://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 15 March 2022)). Wang et al. [13] suggested that higher levels of *CYP2D6* mRNA expression, previously thought to be associated with this SNP, may be explained by LD between rs5758550/rs133333 enhancer SNPs and rs1080985. Haplotypes reconstructed in the studied Croatian Roma population have the rs5758550 allele G and the rs1080985 allele C on more than 20% of chromosomes.

Furthermore, we investigated LDs between enhancer/promotor loci and major star allele-defining SNPs. An r^2 LD value higher than 0.8, indicating a significant association, was found between rs1080983, which is part of the CTCF binding site, and SNP rs16947, which defines allele *2. This SNP is also in high LD with rs28624811 from the promoter/enhancer GH22J042130. SNPs from the same regulatory element, rs1080989 and rs28588594, are in high LD with allele *4 (rs3892097), but an LD value higher than 0.8 was found only in Roma from Balkan, while this high LD value has been observed in all Croatian Roma groups for allele *10 (rs1065852). *CYP2D6*10* is a decreased-function allele

predominantly found in East and South Asian populations, where its prevalence ranges from 9% to 44%. Its frequency in African populations is between 4% and 6%, among Europeans, <2%, and in the Croatian Roma, 6% [24,48,49]. The non-function *CYP2D6*4* allele, which is predominantly found in European populations (18%), had the highest frequency in the Balkan and Baranja Roma groups, even higher than in European populations. The prevalence of the allele *4 in the Medjimurje Roma group is lower than in the European population, but still higher than in other world populations [24,48,49]. Since both these star alleles have an impaired function, an additional analysis such as the reporter gene assay could help to untangle the effect of these promoter SNPs on *CYP2D6* functionality. A high LD was also noticed between the promoter region SNP rs28735595, and SNPs rs1058164 and rs1135840, which are present in all the main *CYP2D6* star alleles. Since the Roma population has a specific genetic history, we were interested to find out whether their *CYP2D6* LD distribution is similar to other world populations. Combinations of the *CYP2D6* regulatory and gene region SNPs with high LD values in the Croatian Roma are also present in the majority of world populations, but fine differences can be noticed among Roma groups. This is especially evident for rs3892097, which defines the *CYP2D6*4* allele, when in LD with rs1080989 and rs28588594 from the promoter region, and this LD is the lowest in the Medjimurje Roma group compared to the other two groups. The LD correlation matrices presented in the MDS plots mostly distinguish the populations of East Asia from other studied populations. Such separation is evident in many studies related to SNPs of the ADME genes [21,50]. Population-specific differences based on r^2 LD values were also found by Ahsan et al. [28], but between drug response-related SNPs.

5. Conclusions

Although the studied Croatian Roma groups showed significant variability of the *CYP2D6* gene variants determined so far, the prevalence of alleles in SNPs from regulatory regions did not differ between these same groups. However, linkage disequilibrium values between these regulatory regions' loci and the *CYP2D6* gene region loci differed between the Croatian Roma groups, and the population of Medjimurje showed the lowest LD values. Higher LD values between the studied SNPs of the promoter region and the SNPs defining impaired-function star alleles *2 and *4 of the *CYP2D6* gene could be used in Roma to improve genotyping efficiency if further studies demonstrate that these promoter SNPs affect the functionality of the CYP2D6 enzyme. An overall comparison of the analyzed LD values revealed that while there was greater variety in the populations of East Asia, they were uniform in populations of Europe and South Asia and distinct in their distribution. In the future, our goal is to sequence the promoter region of the *CYP2D6* gene in Croatian Roma samples, as this would help to further elucidate the structure and frequencies of common overlapping haplotypes of the *CYP2D6* gene, as well as those specific to the Roma population.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm12081353/s1, Figure S1: LD plots for Baranja Roma population, Figure S2: LD plots for Medjimurje Roma population, Figure S3: LD plots for Balkan Roma population.

Author Contributions: Conceptualization, M.P.S. and M.Z.P.; methodology, A.S.M. and M.P.S.; validation, all authors; formal analysis, A.S.M. and M.Z.P.; investigation, A.S.M., M.P.S. and M.Z.P.; data curation, T.Š.-J. and M.Z.P.; writing—original draft preparation, A.S.M. and M.Z.P.; writing—review and editing, all authors; visualization, Ž.C. and M.Š.; supervision, M.P.S.; funding acquisition, M.P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Croatian Science Foundation (IP-2014-09-4454 and DOK-2018-01-4817 to M.P.S.).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Scientific Committee and the Ethics Committee of the Institute for Anthropological Research, in Zagreb, Croatia (RN 1.14-1611/14).

Informed Consent Statement: All Roma participated in the study voluntarily and were informed about the goals, methods and expectations of the study with the help of linguistically and culturally competent and trained Roma volunteers, after which they gave their informed consent.

Data Availability Statement: All data analyzed in this study are available at: http://roma.inantro.hr/en/. In case of using this database for further analyses, please cite this publication. If further clarification is required, contact the corresponding author.

Acknowledgments: We are deeply grateful to the Roma people for their kindness and the interest in participation in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saravanakumar, A.; Sadighi, A.; Ryu, R.; Akhlaghi, F. Physicochemical properties, biotransformation, and transport pathways of established and newly approved medications: A systematic review of the top 200 most prescribed drugs vs. the FDA-approved drugs between 2005 and 2016. *Clin. Pharmacokinet.* **2019**, *58*, 1281–1294. [CrossRef] [PubMed]
2. Hicks, J.K.; Swen, J.J.; Gaedigk, A. Challenges in CYP2D6 phenotype assignment from genotype data: A critical assessment and call for standardization. *Curr. Drug Metab.* **2014**, *15*, 218–232. [CrossRef] [PubMed]
3. Caudle, K.E.; Sangkuhl, K.; Whirl-Carrillo, M.; Swen, J.J.; Haidar, C.E.; Klein, T.E.; Gammal, R.S.; Relling, M.V.; Scott, S.A.; Hertz, D.L.; et al. Standardizing CYP2D6 genotype to phenotype translation: Consensus recommendations from the clinical pharmacogenetics implementation consortium and Dutch pharmacogenetics working group. *Clin. Transl. Sci.* **2019**, *13*, 116–124. [CrossRef] [PubMed]
4. PHARMGKB. Available online: https://www.pharmgkb.org/ (accessed on 15 April 2022).
5. The Human Cythchrome P450 (CYP) Allele Nomenclature Database. Available online: http://cypalleles.ki.se/ (accessed on 15 April 2022).
6. Gaedigk, A.; Simon, S.D.; Pearce, R.E.; Bradford, L.D.; Kennedy, M.J.; Leeder, J.S. The CYP2D6 activity score: Translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther.* **2008**, *83*, 234–242. [CrossRef]
7. Gaedigk, A.; Dinh, J.C.; Jeong, H.; Prasad, B.; Leeder, J.S. Ten years' experience with the CYP2D6 activity score: A perspective on future investigations to improve clinical predictions for precision therapeutics. *J. Pers. Med.* **2018**, *8*, 15. [CrossRef]
8. Fang, Y.; Gao, J.; Wang, T.; Tian, X.; Gao, N.; Zhou, J.; Zhang, H.F.; Wen, Q.; Jin, H.; Xing, Y.R.; et al. Intraindividual variation and correlation of cytochrome P450 activities in human liver microsomes. *Mol. Pharm.* **2018**, *15*, 5312–5318. [CrossRef]
9. Dalton, R.; Lee, S.B.; Claw, K.G.; Prasad, B.; Phillips, B.R.; Shen, D.D.; Wong, L.H.; Fade, M.; McDonald, M.G.; Dunham, M.J.; et al. Interrogation of CYP2D6 structural variant alleles improves the correlation between CYP2D6 genotype and CYP2D6-mediated metabolic activity. *Clin. Transl. Sci.* **2020**, *13*, 147–156. [CrossRef]
10. Ning, M.; Duarte, J.D.; Rubin, L.H.; Jeong, H. CYP2D6 protein level is the major contributor to interindividual variability in CYP2D6-mediated drug metabolism in healthy human liver tissue. *Clin. Pharmacol. Ther.* **2018**, *104*, 974–982. [CrossRef]
11. Pan, X.; Ning, M.; Jeong, H. Transcriptional regulation of CYP2D6 expression. *Drug. Metab. Dispos.* **2017**, *45*, 42–48. [CrossRef]
12. GeneCards (RRID:SCR_002773). Database of Human Genes that Provides Concise Genomic, Proteomic, Transcriptomic, Genetic and Functional Information on All Known and Predicted Human Genes. Information featured in Genecards Includes Orthologies, Disease Relationships, Mutations And SNPS, Gene Expression, Gene Function, Pathways, Protein-Protein Interactions, Related Drugs and Com-pounds and Direct Links to Cutting Edge Research Reagents and Tools Such as Antibodies, Recombinant Proteins, Clones, Expression Assays and RNAi Reagents. Available online: http://genecards.org (accessed on 15 April 2022).
13. Wang, D.; Poi, M.J.; Sun, X.; Gaedigk, A.; Leeder, J.S.; Sadee, W. Common CYP2D6 polymorphisms affecting alternative splicing and transcription: Long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum. Mol. Genet.* **2014**, *23*, 268–278. [CrossRef]
14. Gaedigk, A.; Casey, S.T.; Whirl-Carrillo, M.; Miller, N.A.; Klein, T.E. Pharmacogene variation consortium: A global resource and repository for pharmacogene variation. *Clin. Pharmacol. Ther.* **2021**, *110*, 542–545. [CrossRef] [PubMed]
15. Gaedigk, A.; Ingelman-Sundberg, M.; Miller, N.A.; Leeder, J.S.; Whirl-Carrillo, M.; Klein, T.E. The pharmacogene variation (PharmVar) consortium: Incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin. Pharmacol. Ther.* **2018**, *103*, 399–401. [CrossRef] [PubMed]
16. Gaedigk, A.; Whirl-Carrillo, M.; Pratt, V.M.; Miller, N.A.; Klein, T.E. PharmVar and the landscape of pharmacogenetic resources. *Clin. Pharmacol. Ther.* **2020**, *107*, 43–46. [CrossRef] [PubMed]
17. Gong, X.; Liu, Y.; Zhang, X.; Wei, Z.; Huo, R.; Shen, L.; He, L.; Qin, S. Systematic functional study of cytochrome P450 2D6 promoter polymorphisms in the Chinese Han population. *PLoS ONE* **2013**, *8*, e57764. [CrossRef]
18. Fraser, A. *The Gypsies*; Blackwell Publishers: Oxford, UK, 1992.
19. Gresham, D.; Morar, B.; Underhill, P.A.; Passarino, G.; Lin, A.A.; Wise, C.; Angelicheva, D.; Calafell, F.; Oefner, P.J.; Shen, P.; et al. Origins and divergence of the Roma (gypsies). *Am. J. Hum. Genet.* **2001**, *69*, 1314–1331. [CrossRef]
20. Chaix, R.; Austerlitz, F.; Morar, B.; Kalaydjieva, L.; Heyer, E. Vlax Roma history: What do coalescent-based methods tell us? *Eur. J. Hum. Genet.* **2004**, *12*, 285–292. [CrossRef]

21. Škarić-Jurić, T.; Tomas, Ž.; Zajc Petranović, M.; Božina, N.; Smolej Narančić, N.; Janićijević, B.; Salihović, M.P. Characterization of ADME genes variation in Roma and 20 populations worldwide. *PLoS ONE* **2018**, *13*, e0207671. [CrossRef]
22. Miller, S.A.; Dykes, D.D.; Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **1988**, *16*, 1215. [CrossRef]
23. Semagn, K.; Babu, R.; Hearne, S.; Olsen, M. Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* **2014**, *33*, 1–14. [CrossRef]
24. Stojanović Marković, A.; Zajc Petranović, M.; Tomas, Ž.; Puljko, B.; Šetinc, M.; Škarić-Jurić, T.; Peričić Salihović, M. Untangling SNP variations within CYP2D6 gene in Croatian Roma. *J. Pers. Med.* **2022**, *12*, 374. [CrossRef]
25. Excoffier, L.; Lischer, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [CrossRef]
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: https://www.R-project.org/ (accessed on 22 January 2022).
27. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [CrossRef]
28. Ahsan, T.; Urmi, N.J.; Sajib, A.A. Heterogeneity in the distribution of 159 drug-response related SNPs in world populations and their genetic relatedness. *PLoS ONE* **2020**, *15*, e0228000. [CrossRef] [PubMed]
29. Evans, D.M.; Cardon, L.R. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *Am. J. Hum. Genet.* **2005**, *76*, 681–687. [CrossRef]
30. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [CrossRef] [PubMed]
31. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; Available online: https://ggplot2.tidyverse.org (accessed on 25 January 2022).
32. Nagaraj, S.H.; Toombs, M. The gene-drug duality: Exploring the pharmacogenomics of indigenous populations. *Front. Genet.* **2021**, *12*, 687116. [CrossRef]
33. Font-Porterias, N.; Giménez, A.; Carballo-Mesa, A.; Calafell, F.; Comas, D. Admixture has shaped romani genetic diversity in clinically relevant variants. *Front. Genet.* **2021**, *12*, 683880. [CrossRef] [PubMed]
34. Zajc Petranovic, M.; Tomas, Z.; Skaric-Juric, T.; Smolej Narancic, N.; Janicijevic, B.; Stojanovic Markovic, A.; Pericic Salihovic, M. The variability of multi-drug resistance ABCB1 gene in the Roma population from Croatia. *Mol. Exp. Biol. Med.* **2019**, *2*, 10–18. [CrossRef]
35. Sipeky, C.; Csongei, V.; Jaromi, L.; Safrany, E.; Maasz, A.; Takacs, I.; Beres, J.; Fodor, L.; Szabo, M.; Melegh, B. Genetic variability and haplotype profile of MDR1 (ABCB1) in Roma and Hungarian population samples with a review of the literature. *Drug Metab. Pharmacokinet.* **2011**, *26*, 206–215. [CrossRef]
36. Tomas, Z.; Kuhanec, A.; Skaric-Juric, T.; Zajc Petranovic, M.; Smolej Narancic, N.; Janicijevic, B.; Pericic Salihovic, M. Distinctiveness of the Roma population within CYP2B6 worldwide variation. *Pharmacogenomics* **2017**, *18*, 1575–1587. [CrossRef]
37. Weber, A.; Szalai, R.; Sipeky, C.; Magyari, L.; Melegh, M.; Jaromi, L.; Matyas, P.; Duga, B.; Kovesdi, E.; Hadzsiev, K.; et al. Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples. *Pharmacol. Rep.* **2015**, *67*, 460–464. [CrossRef] [PubMed]
38. Dlouhá, L.; Adámková, V.; Šedová, L.; Olišarová, V.; Hubáček, J.A.; Tóthová, V. Five genetic polymorphisms of cytochrome P450 enzymes in the Czech non-Roma and Czech Roma population samples. *Drug Metab. Pers. Ther.* **2020**, *35*, 20200103. [CrossRef] [PubMed]
39. Zajc Petranovic, M.; Tomas, Z.; Skaric-Juric, T.; Smolej Narancic, N.; Janicicjevic, B.; Pericic Salihovic, M. The variation of CYP2C19 gene in the Roma population from Croatia. *Mol. Exp. Biol. Med.* **2018**, *1*, 32–37.
40. Sipeky, C.; Weber, A.; Szabo, M.; Melegh, B.I.; Janicsek, I.; Tarlos, G.; Szabo, I.; Sumegi, K.; Melegh, B. High prevalence of CYP2C19*2 allele in Roma samples: Study on Roma and Hungarian population samples with review of the literature. *Mol. Biol. Rep.* **2013**, *40*, 4727–4735. [CrossRef]
41. Petrović, J.; Pešić, V.; Lauschke, V.M. Frequencies of clinically important CYP2C19 and CYP2D6 alleles are graded across Europe. *Eur. J. Hum. Genet.* **2019**, *28*, 88–94. [CrossRef]
42. Teixeira, J.; Amorim, A.; Prata, M.J.; Quental, S. Pharmacogenetic polymorphisms in a Portuguese gypsy population. *Curr. Pharm. Person. Med.* **2015**, *13*, 36–40. [CrossRef]
43. Stojanović Marković, A.; Zajc Petranović, M.; Škobalj, M.; Poloni, E.S.; Pichler Oberški, L.; Škarić-Jurić, T.; Pericic Salihović, M. From dietary adaptation in the past to drug metabolism of today: An example of NAT genes in the Croatian Roma. *Am. J. Biol. Anthropol.* **2022**, *178*, 140–153. [CrossRef]
44. Elias, A.B.R.; Araújo, G.S.; de Souza, S.J.; Suarez-Kurtz, G. Distribution and linkage disequilibrium of the enhancer SNP rs5758550 among Latin American populations: Influence of continental ancestry. *Pharm. Genom.* **2020**, *30*, 67–72. [CrossRef]
45. Wang, D.; Papp, A.C.; Sun, X. Functional characterization of CYP2D6 enhancer polymorphisms. *Hum. Mol. Genet.* **2015**, *24*, 1556–1562. [CrossRef]
46. Raimundo, S.; Fischer, J.; Eichelbaum, M.; Griese, E.U.; Schwab, M.; Zanger, U.M. Elucidation of the genetic basis of the common 'intermediate metabolizer' phenotype for drug oxidation by CYP2D6. *Pharmacogenetics* **2000**, *10*, 577–581. [CrossRef]

47. Zanger, U.M.; Fischer, J.; Raimundo, S.; Stüven, T.; Evert, B.O.; Schwab, M.; Eichelbaum, M. Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics* **2011**, *11*, 573–585. [CrossRef] [PubMed]
48. Pratt, V.M.; Cavallari, L.H.; Del Tredici, A.L.; Gaedigk, A.; Hachad, H.; Ji, Y.; Kalman, L.V.; Ly, R.C.; Moyer, A.M.; Scott, S.A.; et al. Recommendations for clinical CYP2D6 genotyping allele selection. *J. Mol. Diagn.* **2021**, *23*, 1047–1064. [CrossRef] [PubMed]
49. Gaedigk, A.; Sangkuhl, K.; Whirl-Carrillo, M.; Klein, T.; Leeder, J.S. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* **2017**, *19*, 69–76. [CrossRef] [PubMed]
50. Li, J.; Zhang, L.; Zhou, H.; Stoneking, M.; Tang, K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* **2011**, *20*, 528–540. [CrossRef]

Figure S1: LD plots for Baranja Roma population

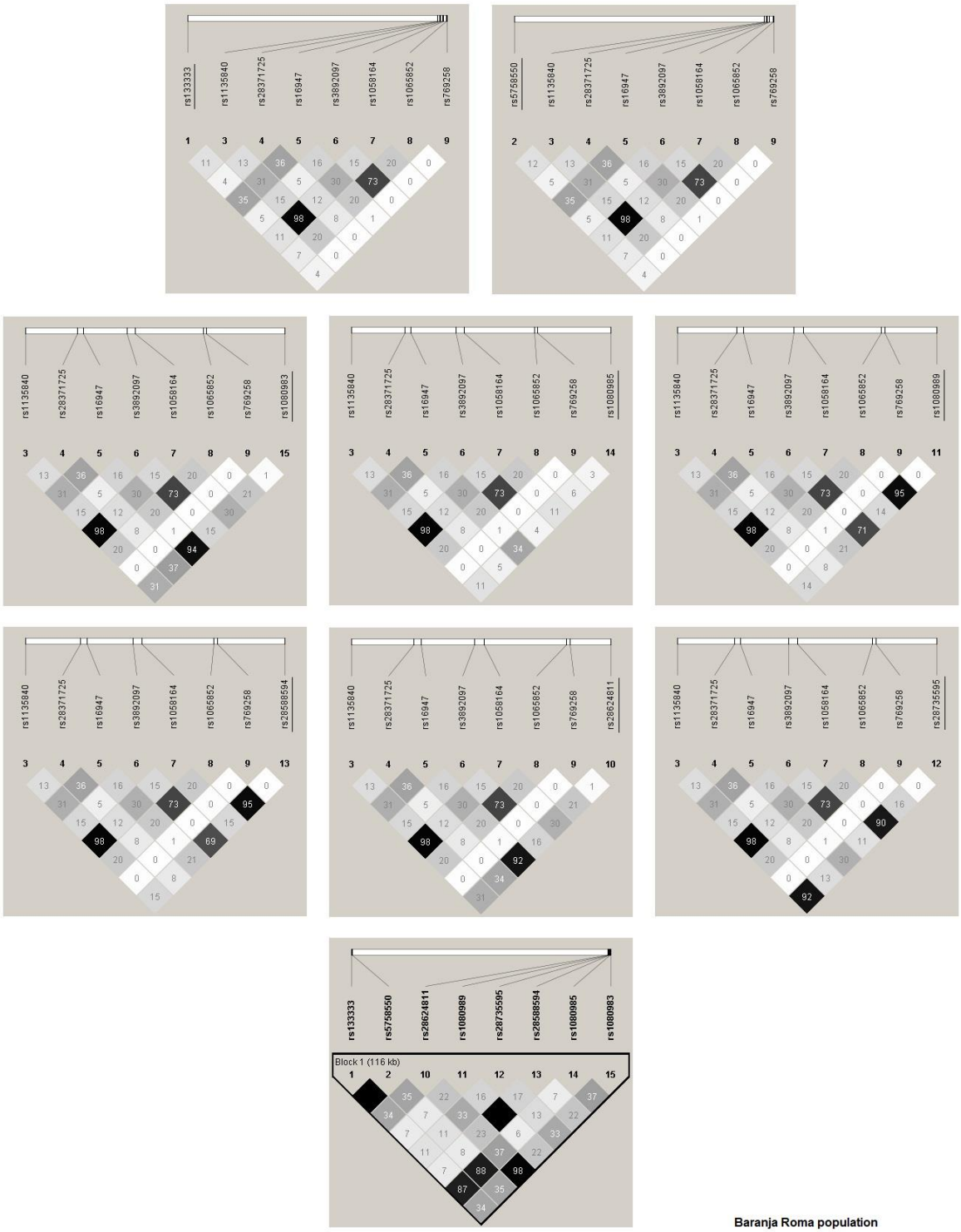


Figure S2: LD plots for Medjimurje Roma population

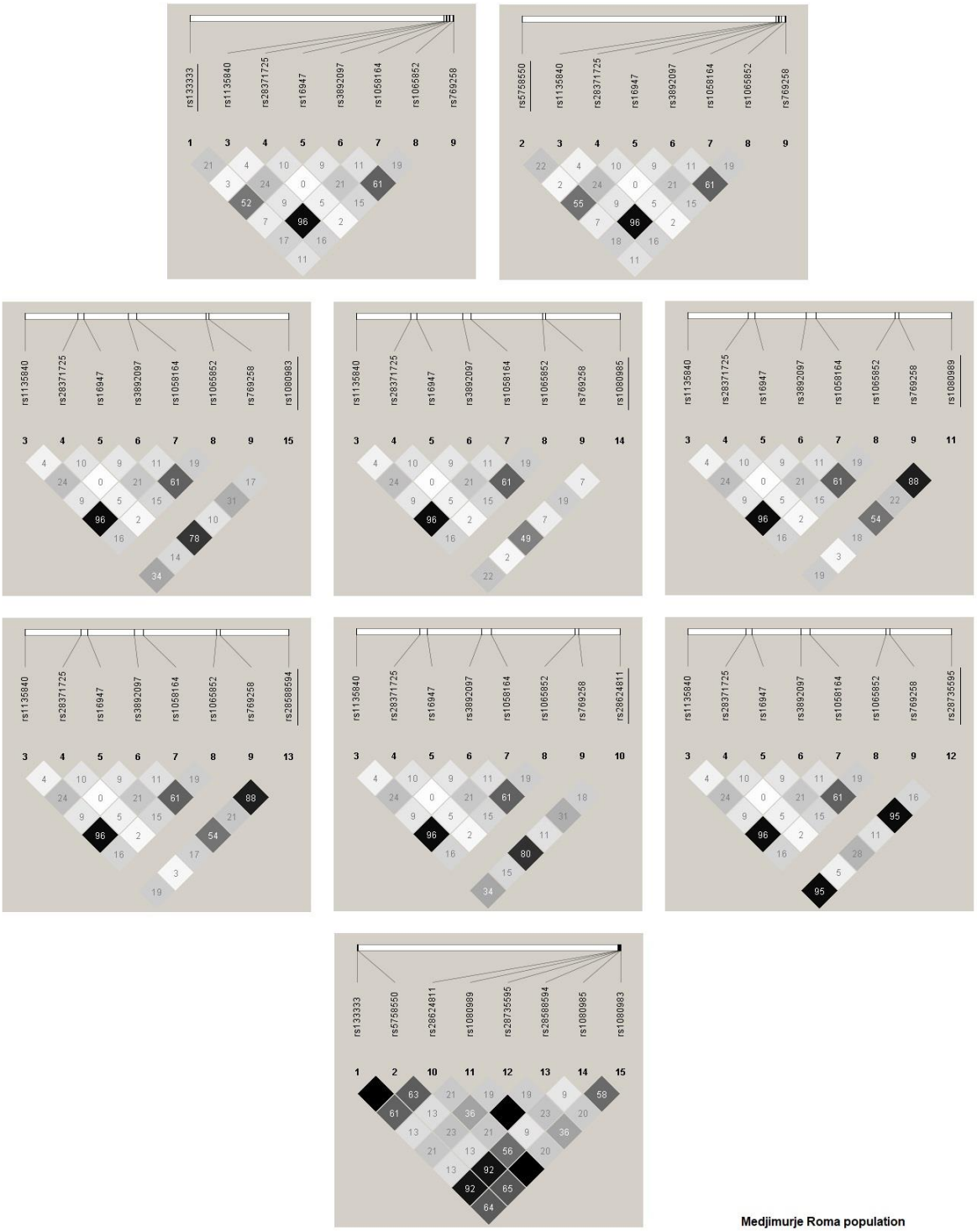
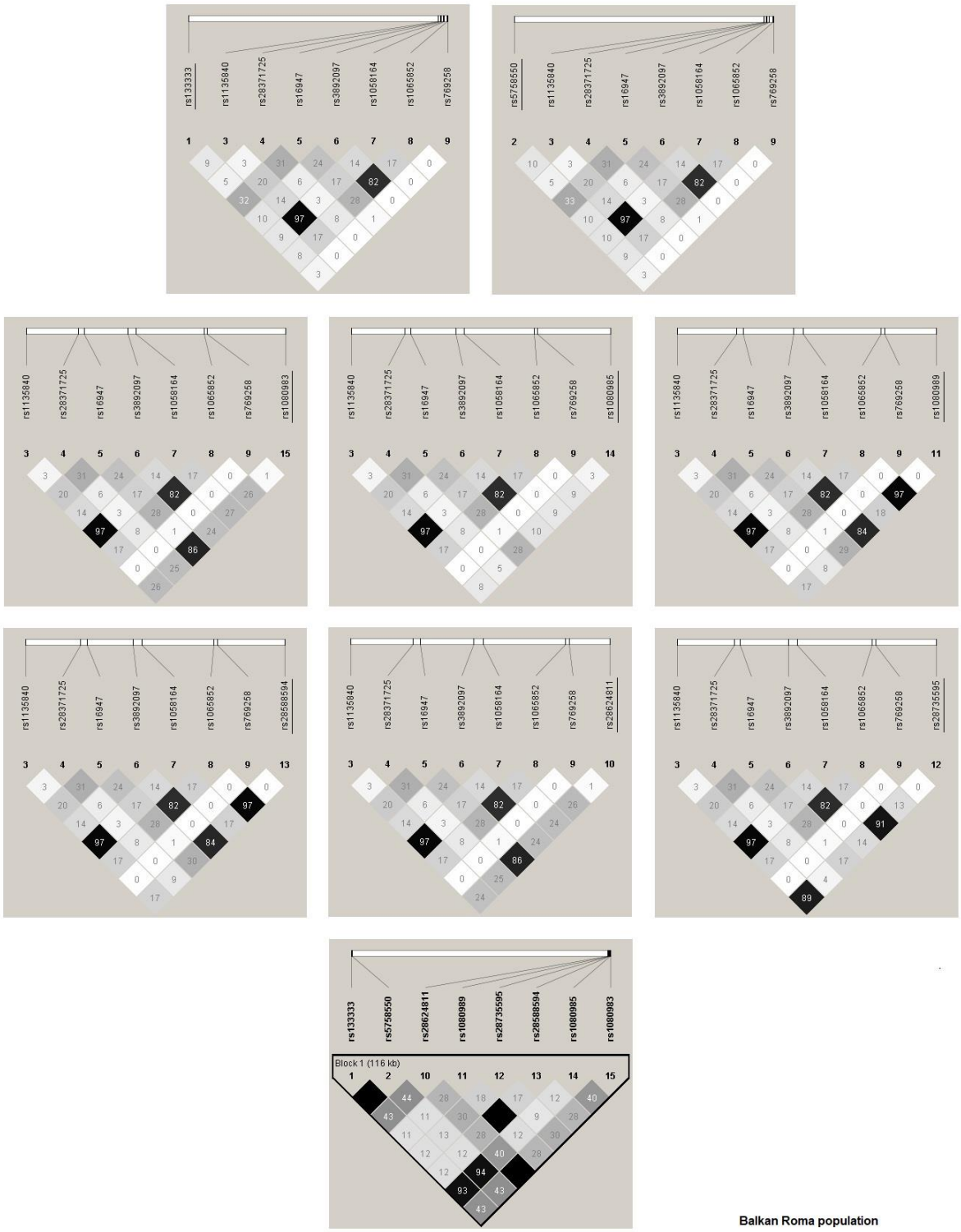


Figure S3: LD plots for Balkan Roma population



2.3. RELEVANCE OF *CYP2D6* GENE VARIANTS IN POPULATION GENETIC DIFFERENTIATION

pharmaceutics

MDPI

Article

Relevance of *CYP2D6* Gene Variants in Population Genetic Differentiation

Anita Stojanović Marković [1], Matea Zajc Petranović [1], Tatjana Škarić-Jurić [1], Željka Celinšćak [1], Maja Šetinc [1], Željka Tomas [2] and Marijana Peričić Salihović [1,*]

[1] Institute for Anthropological Research, 10000 Zagreb, Croatia
[2] Department for Translational Medicine, Srebrnjak Children's Hospital, 10000 Zagreb, Croatia
* Correspondence: mpericic@inantro.hr

Abstract: A significant portion of the variability in complex features, such as drug response, is likely caused by human genetic diversity. One of the highly polymorphic pharmacogenes is *CYP2D6*, encoding an enzyme involved in the metabolism of about 25% of commonly prescribed drugs. In a directed search of the 1000 Genomes Phase III variation data, 86 single nucleotide polymorphisms (SNPs) in the *CYP2D6* gene were extracted from the genotypes of 2504 individuals from 26 populations, and then used to reconstruct haplotypes. Analyses were performed using Haploview, Phase, and Arlequin softwares. Haplotype and nucleotide diversity were high in all populations, but highest in populations of African ancestry. Pairwise F_{ST} showed significant results for eleven SNPs, six of which were characteristic of African populations, while four SNPs were most common in East Asian populations. A principal component analysis of *CYP2D6* haplotypes showed that African populations form one cluster, Asian populations form another cluster with East and South Asian populations separated, while European populations form the third cluster. Linkage disequilibrium showed that all African populations have three or more haplotype blocks within the *CYP2D6* gene, while other world populations have one, except for Chinese Dai and Punjabi in Pakistan populations, which have two.

Keywords: *CYP2D6* gene; polymorphism; haplotype; star allele; pharmacogenetics; 1000 Genomes populations

Citation: Stojanović Marković, A.; Zajc Petranović, M.; Škarić-Jurić, T.; Celinšćak, Ž.; Šetinc, M.; Tomas, Ž.; Peričić Salihović, M. Relevance of *CYP2D6* Gene Variants in Population Genetic Differentiation. *Pharmaceutics* **2022**, *14*, 2481. https://doi.org/10.3390/pharmaceutics14112481

Academic Editors: Rocio Nuñez-Torres and Anna González-Neira

Received: 13 October 2022
Accepted: 10 November 2022
Published: 16 November 2022

1. Introduction

Human CYP2D6 protein was purified in 1984 [1], and the gene was mapped to chromosome 22q13 in 1987 [2]. Two years later, the *CYP2D6* gene was cloned and sequenced [3], and it was discovered that the gene locus contains two additional genes: a non-functional *CYP2D7* gene, and a *CYP2D8* pseudogene. The *CYP2D6* gene contains nine exons and is highly polymorphic.

In 1996, a group of international experts in pharmacogenetics decided to systematize allelic variants of the *CYP2D6* by proposing a haplotype-based star (*) nomenclature system [4]. Since then, more than 135 *CYP2D6* star alleles have been described and are available on the Pharmacogene Variation (PharmVar) Consortium website [5,6].

Genetic variations of the *CYP2D6* affect the metabolizing activity of the CYP2D6 enzyme. Its activity can vary from complete absence to increased activity, and can be grouped into four different drug-metabolism phenotypes: (1) poor metabolizer (PM—two null activity alleles); (2) intermediate metabolizer (IM—one normal activity allele with one null activity allele; or two reduced activity alleles); (3) extensive metabolizer (EM—two normal activity alleles; or a combination of one increased activity allele with one allele of reduced activity); and (4) ultra-rapid metabolizer (UM—one normal activity allele with one increased activity allele) [7–10]. CYP2D6 is expressed in the human liver where it accounts for only 2–4% of the total CYP content [11,12], but it is involved in the metabolism of up to

25% of drugs commonly used in medicine, including antidepressants, a number of atypical and typical antipsychotics, antineoplastic agents (e.g., tamoxifen), adrenergic antagonists (e.g., metoprolol), and analgesics (e.g., codeine and tramadol) [13–19]. Variations in the *CYP2D6* gene have also been studied as a risk factor for a number of diseases: Parkinson's disease [20–22], schizophrenia and other psychiatric diseases [16,23], Alzheimer's disease [24,25], as well as several forms of cancer [26,27].

Although CYP2D6′s role in the metabolism of naturally occurring xenobiotics has not been researched extensively, it is well-known that this enzyme has a very high affinity for alkaloids [28]. Therefore, alkaloid metabolism in food is assumed to have played a role in its evolution. There is a theory that 10,000 to 20,000 years ago in Northwest Africa, due to food shortages compared to population size, the number of plants that could provide usable food increased as a result of selection that favored the survival of individuals capable of more effective detoxification of plant toxins [29]. The best example of how dietary modifications throughout human history have provoked selection pressure on the genes whose products metabolize food molecules is N-acetyl-transferase 2 (e.g., [30]). The current patterns of *CYP2D6* genetic diversity, according to Fuselli (2010), are a result of the selective pressure of new or more potent CYP2D6 substrates that emerged as food choices, particularly at the start of the Neolithic transition, in response to worsening nutritional conditions and higher disease burdens [31].

More genetic variation can be seen in genes encoding detoxification enzymes, which shows that exposure to various substrates also aided in the evolution of genetic variants. Detoxification enzymes really exhibit signs of positive selection, such as modifications of the amino acid sequence that increase substrate selectivity [6,32]. Compared to any other category of pharmacogenomically relevant genes in humans, a recent study revealed that *CYP* genes that metabolize exogenous compounds have far higher frequencies of SNPs that vary greatly between populations [33].

It has not yet been possible to pinpoint the selective factor that causes diet-related patterns of evolution in the *CYP2D6* gene (such as the presence or absence of a particular substrate or a variable concentration of substrates) [34].

Although there are numerous papers on the world distribution of *CYP2D6* variation, most focus on the pharmacological consequences of different variants. The goal of this paper was to identify which single nucleotide polymorphisms (SNPs) and haplotypes in the *CYP2D6* gene determine the genetic specificity of 26 world populations, and to test intra- and inter-group differences in continental groups defined by ancestry. Furthermore, we investigated the role of population differentiation in the definition of the *CYP2D6* star alleles.

2. Materials and Methods

The investigated pharmacogene *CYP2D6* is located on the reverse strand of chromosome 22:42,126,499–42,130,865 (GRCh38). Using Data Slicer, a tool implemented on the Ensembl website [35], data on 2504 individuals belonging to 26 world populations from Phase 3 of the 1000 Genomes Project were extracted from Ensembl Release 107 [36]. The data file contained 279 polymorphic positions: 9 insertions/deletions (indels) and 270 single nucleotide polymorphisms (SNPs). Two positions were monomorphic, and in 182 positions the minor allele was found less than five times in the total sample. All indels, monomorphic SNPs, and SNPs where the minor allele occurred less than five times in the total sample were excluded from further analyses, leaving 86 SNPs.

Allele frequencies and the Hardy–Weinberg equilibrium were calculated separately for each population using VCFtools [37]. VCF files were also used to create ped files for linkage disequilibrium (LD) calculation and visualization, which was performed in Haploview software [38]. Haplotype blocks were constructed using the confidence intervals algorithm [39], and the informativeness of the block was further estimated by r^2 measurement of LD between SNPs defining the ends of haplotype blocks, both implemented in Haploview software. The *CYP2D6* haplotypes were inferred using Phase ver. 2.1 [40,41]. Haplotype

frequencies were used for principal component analysis (PCA), performed using the statistical package SPSS Statistics 21.0 for Windows (SPSS Inc., Chicago, IL, USA). The most common haplotypes were translated into the star allele nomenclature using data on the PharmVar website [5].

Indices of intrapopulation molecular diversity (number of haplotypes, polymorphic sites, transitions and transversions) and AMOVA, the statistical significance of which was assessed by generating 100,000 random samples, were calculated using Arlequin 3.5 software [42].

3. Results

In order to capture most of the variability in the *CYP2D6* gene, we investigated 86 SNPs within the gene region (22:42,126,499–42,130,865), whose minor allele frequencies (MAF) were higher than 0.1%. The allele frequencies of studied polymorphic sites in 26 world populations from the 1000 Genomes database are shown in Supplementary Table S1. Phased SNPs revealed 232 unique haplotypes.

Several diversity indices were calculated in order to see the intrapopulation variation, and the findings are displayed in Table 1.

African populations had the highest number of polymorphic sites: the most polymorphisms (55) were found in the population of African ancestry in the Southwest USA, and the least (46) in the Mende population in Sierra Leone. The lowest number of polymorphic sites was found in East Asian populations ranging from 20 in the Japanese population to 30 in the Han Chinese population in Beijing and in the Kinh population in Vietnam. The highest number of haplotypes were found in populations of African ancestry, ranging from 37 in the Gambian population to 31 haplotypes in the Yoruba population from Nigeria. The lowest number of haplotypes of all 26 investigated populations was found in the population of Japan (11), followed by the population of Peru (16) and the British population from England and Scotland (17). Interestingly, although considered a genetically isolated population, Finns did not have the lowest number of haplotypes.

Table 1. The diversity indices in 26 world populations from the 1000 Genomes database.

		Sample Size	No. of Haplotypes	No. of Polymorphic Sites	Haplotype Diversity	Nucleotide Diversity
African Ancestry	**ESN**	198	36	50	0.935	0.113
	GWD	226	37	49	0.918	0.114
	MSL	170	32	46	0.933	0.113
	YRI	216	31	48	0.926	0.109
	LWK	198	35	48	0.936	0.113
	ACB	192	36	48	0.933	0.115
	ASW	122	36	55	0.908	0.109
American Ancestry	**CLM**	188	26	44	0.813	0.106
	MXL	128	23	36	0.764	0.092
	PEL	170	16	34	0.602	0.086
	PUR	208	32	56	0.875	0.113
European Ancestry	**CEU**	198	21	38	0.847	0.112
	FIN	198	19	36	0.778	0.102
	GBR	182	17	35	0.834	0.112
	IBS	214	27	47	0.851	0.110
	TSI	214	22	41	0.845	0.113

Table 1. *Cont.*

		Sample Size	No. of Haplotypes	No. of Polymorphic Sites	Haplotype Diversity	Nucleotide Diversity
South Asian Ancestry	**BEB**	172	22	34	0.809	0.099
	GIH	206	22	36	0.787	0.106
	ITU	204	21	34	0.794	0.105
	PJL	192	18	37	0.713	0.094
	STU	204	28	38	0.784	0.102
East Asian Ancestry	**CDX**	186	18	26	0.620	0.069
	CHB	206	19	30	0.637	0.075
	CHS	210	20	27	0.643	0.071
	JPT	208	11	23	0.717	0.074
	KHV	198	22	30	0.607	0.066

Population abbreviations: European ancestry: CEU (Utah residents with Northern and Western ancestry), FIN (Finland), GBR (British in England and Scotland), IBS (Iberian population in Spain), TSI (Toscani in Italy); South Asian ancestry: BEB (Bengali in Bangladesh), GIH (Gujarati Indian), ITU (Indian Telugu in the UK), PJL (Punjabi in Lahore Pakistan), STU (Sri Lankan Tamil in the UK); African ancestry: ACB (African Caribbean in Barbados), ASW (African Ancestry in South West USA), ESN (Esan in Nigeria), GWD (Gambian in Western Division), LWK (Luhya in Webuye, Kenya), MSL (Mende in Sierra Leone), YRI (Yoruba in Ibadan, Nigeria); American ancestry: CLM (Colombian in Medellin, Colombia), MXL (Mexico), PEL (Peruvian in Lima, Peru), PUR (Puerto Rican in Puerto Rico); East Asian ancestry: CDX (Dai Chinese), CHB (Han Chinese in Beijing), CHS (Southern Han Chinese), JPT (Japanese in Tokyo), KHV (Kinh in Ho Chi Minh City, Vietnam).

Overall, haplotype diversity and nucleotide diversity were high in all populations, and highest in African populations where haplotype diversity ranged from 0.908 to 0.936. The lowest haplotype diversity was observed in East Asian populations ranging from 0.607 (Vietnam) to 0.717 (Japan). The results of the nucleotide diversity analysis are also very similar: the highest diversity was found in African populations, and the lowest in East Asian populations. According to diversity indices, the Japanese population has the lowest genetic variation.

In order to calculate the level of population differentiation, we performed AMOVA analyses. Populations were joined in five continental groups based on shared common ancestry. Approximately 8% of the variation was due to between-group differences (F_{CT} = 0.077), while the interpopulation variation was 9% (F_{ST} = 0.091). When we examined each continental group separately, we discovered that East Asian populations showed the greatest differentiation (F_{ST} = 0.031), while European populations showed the smallest (F_{ST} = 0.002).

To elucidate which of the 86 *CYP2D6* gene SNPs mostly affected the population differentiation, locus-by-locus AMOVA was conducted. Allelic variations at 11 SNPs (rs75203276, rs59421388, rs61736512, rs16947, rs76327133, rs80262685, rs28371706, rs2267447, rs1065852, rs2004511, and rs1081003) contribute to the inter-population differentiation higher than 10%, with F_{ST} values ranging from 0.103 to 0.366 (Table 2).

Figure 1 shows the distribution of minor alleles in those 11 SNPs in 26 world populations. Six SNPs are characteristic of African populations (rs75203276, rs59421388, rs61736512, rs76327133, rs80262685, and rs28371706), with the population of African ancestry in Southwest USA showing somewhat lower frequencies with values of 3–4% in five SNPs. The remaining five SNPs are present in all world populations, with rs16947 being most common in African populations (42–65%), least common in East Asian populations (13–17%), and occurring in a range of 25–44% in the rest of the world population. The SNPs rs2267447, rs1065852, rs2004511, and rs1081003 are the most common in East Asian populations with frequencies in the range of 60–68%, while in the Japanese population they were in the range of 36–39%. The most common SNPs for non-African and non-Asian populations are rs16947, rs226744, rs106585, and rs200451.

Table 2. List of 11 single nucleotide polymorphisms (SNPs) of the *CYP2D6* gene which show inter-population differentiation (F_{ST}) above 0.1. F_{CT} defines the proportion of total genetic variability among continental groups.

SNP	Position	F_{ST}	*p*	F_{CT}	*p*
rs1081003	42129754	0.366	<0.00001	0.343	0.00001
rs2004511	42127209	0.211	<0.00001	0.193	0.00001
rs1065852	42130692	0.209	<0.00001	0.188	<0.00001
rs2267447	42128694	0.204	<0.00001	0.187	<0.00001
rs28371706	42129770	0.197	<0.00001	0.163	0.0001
rs80262685	42128576	0.117	<0.00001	0.096	0.00015
rs76327133	42128668	0.115	<0.00001	0.094	0.00025
rs16947	42127941	0.111	<0.00001	0.099	<0.00001
rs61736512	42129132	0.105	<0.00001	0.085	0.00026
rs59421388	42127608	0.104	<0.00001	0.084	0.0002
rs75203276	42128499	0.103	<0.00001	0.081	0.00014

The distribution of haplotypes in all 26 investigated populations is shown in Figure 2. Haplotypes 1 and 2 are present in all world populations with varying frequencies, and their combined frequencies range from a minimum of 16% in the population of Sierra Leone to a maximum of 82% in the Peru population. Haplotype 3 is distributed in all populations except the Finnish population. However, it is most characteristic of East Asian populations, with the lowest distribution of 36% in Japan, while other East Asian populations have it in the range of 56–60%. This haplotype is found in some African populations with a frequency of more than 5% (Nigeria, the African Caribbean and Gambia) and higher than 10% in Sierra Leone and Bangladesh. The distribution frequencies of haplotype 4 are over 10% in all European populations (minimum 12% in Finland, maximum 19% in Great Britain), and in populations of Puerto Rico and Colombia. Haplotype 5 is most common in Southeast Asian populations, but has a frequency distribution of over 10% in Italian and Central European populations. Haplotypes 6, 7, 8, and 10 are mainly characteristic of African populations, where the frequency of distribution is generally higher than 5%. All haplotypes which occurred less than five times were presented together (rest) in Figure 2.

All populations share the most common haplotype determining the star allele *1. The eight most common haplotypes account for 74% of all haplotypes worldwide. Haplotype 11 translated into star allele *10 is typical for East Asian populations. The most common haplotypes in Europe are *2 and *4. In South Asian populations, the most common haplotypes are *2 and *41. African populations are the most specific: star alleles *1, *29, and *17 are predominantly found in these populations.

To compare world populations based on *CYP2D6* haplotypes, we performed a principal component analysis (PCA) (Figure 3). Its results showed that African populations form one cluster, Asian populations another cluster with East and South Asian populations separated, and European populations form a third cluster. South American populations do not have a distinct cluster: Colombian and Puerto Rican populations overlap with European populations, while Mexican and Peruvian populations are closer to Southeast Asian populations.

Pharmaceutics **2022**, *14*, 2481

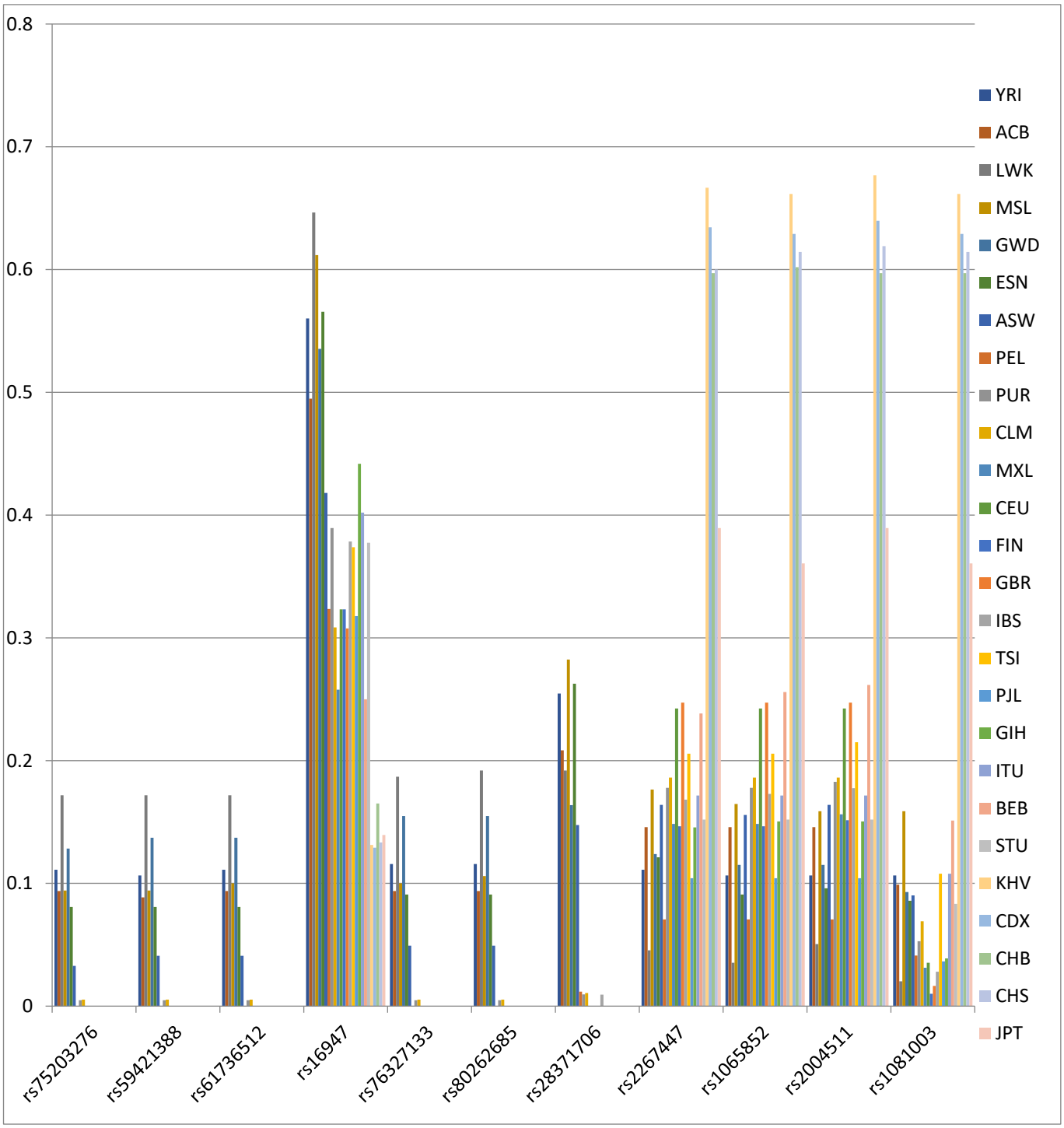


Figure 1. Distribution of minor alleles frequencies in 11 SNPs of the *CYP2D6* gene in 26 world populations.

Pharmaceutics **2022**, *14*, 2481

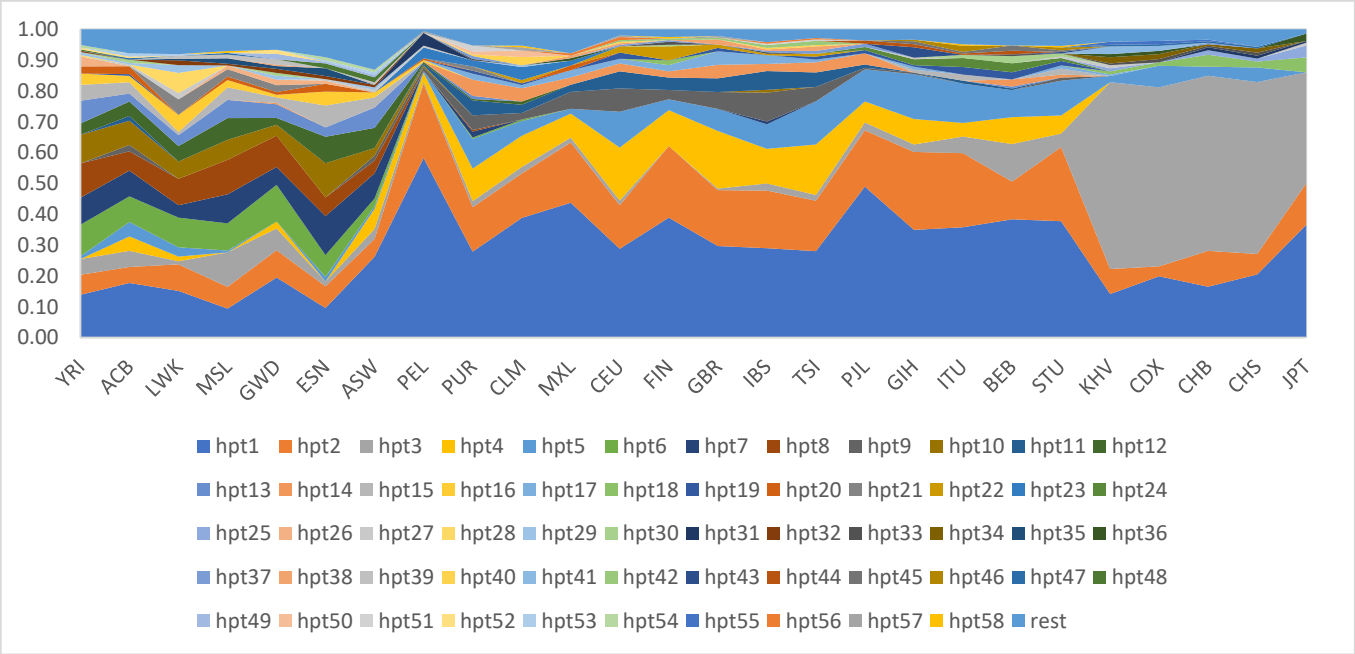


Figure 2. Distribution of the *CYP2D6* gene haplotypes in 26 world populations.

Linkage disequilibrium (LD) was calculated and visualized using Haploview 4.2 software, which constructs blocks based on D′ values. All African populations have three or more haplotype blocks within the *CYP2D6* gene, while other world populations have one haplotype block, except Chinese Dai and Punjabi in Pakistan populations, which have two. A large block of the Chinese Dai population (in the range rs1135840–rs28371702) has an r^2 of 0.888 and a large block of the Punjabi population in Pakistan (in the range rs16947–rs1080995) has an r^2 of 0.976, while both small blocks in these two populations have an r^2 below 0.8. In all African populations but Yoruba from Nigeria, we observed the *CYP2D6* gene haplotype block ranging from rs1081000 to rs1080995 ($r^2 > 0.8$). All African populations except the African ancestry in the southwest USA population share the same haplotype block in the range rs1135840–rs27371730 (with r^2 substantially below 0.8), and ASW is the only African population to have all its blocks with r^2 greater than 0.8. The Yoruba in Ibadan (Nigeria) population has a total of four haplotype blocks, but only one block is in complete LD (in the range rs75203276–rs61736512). The same haplotype block occurs in the Mende in Sierra Leone population (r^2 of 0.935) and the African Caribbean in Barbados population (r^2 of 1.0). In addition to the block in the range rs1081000–1065850, the Luhya in Kenya population has another one that is in high LD (in the range rs75203276–rs76327133, r^2 of 0.902). The Gambian in Western Division-Mandinka population's second haplotype block in high LD is the one in range rs569421388–rs76327133, with an r^2 of 0.868.

The haplotype block found In the European continental group, ranging from rs1135840 to rs1065852, in all five populations has r^2 values substantially below 0.8. The same haplotype block was also detected in three East Asian populations (Japanese and two Han Chinese populations), in three South Asian populations (Indian Telugu, Bengali in Bangladesh, and Sri Lankan Tamil in the UK), and in two American populations (Colombian and Puerto Rican), also with r^2 below 0.8. The only remaining East Asian population, Kinh from Vietnam, has a haplotype block bit shorter than the one previously mentioned, ranging from rs28371730 to rs1065852. In contrast, the population of Gujarati Indians in the USA, the last remaining South Asian population, also has a shorter block ranging from rs1135840 to rs1080995, with both r^2 below 0.8. The latter block, again with $r^2 < 0.8$, was also found in the Mexican Ancestry in Los Angeles and Peruvian populations.

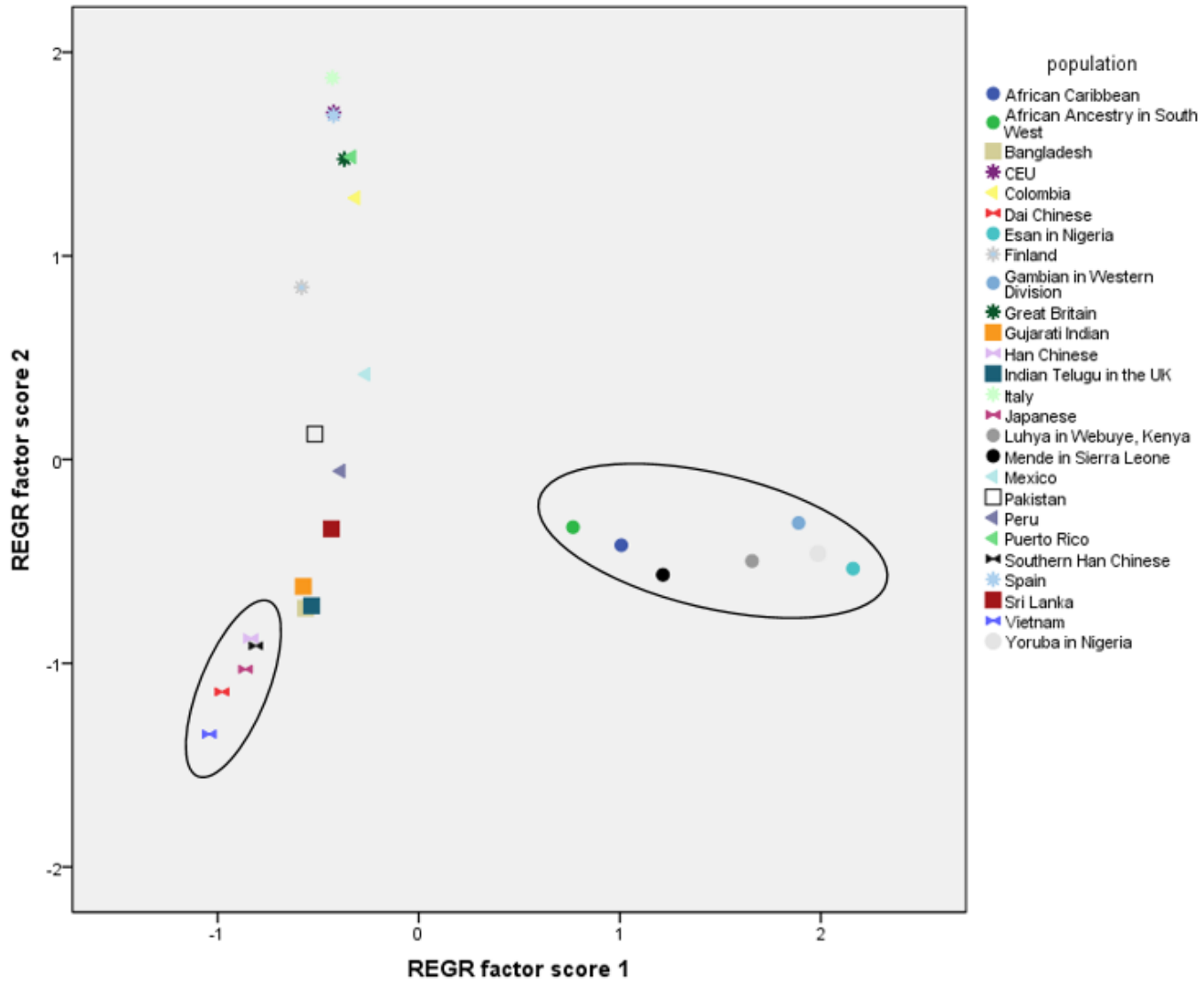


Figure 3. Principal component analysis (PCA) based on *CYP2D6* haplotypes present in 26 world populations.

4. Discussion

Genetic polymorphisms are responsible for a substantial proportion of inter-individual and inter-ethnic heterogeneity in drug response [43]. A number of studies investigated the distribution of genetic variants responsible for heterogeneity in drug response in different populations [44–46]. In this study, 86 polymorphic SNPs within the *CYP2D6* gene were analysed in 26 world populations from the 1000 Genomes database [36], in order to estimate the influence of *CYP2D6* haplotype diversity on population differentiation.

The estimated F_{ST} (0.07) indicates a moderate overall genetic differentiation level. However, when continental population groups were examined separately, European populations showed more than 10 times lower interpopulation differentiaton than East Asian populations, which showed the highest. Similar results were found in a study by Jay and colleagues (2013), which showed that the lowest genetic variation of the ADME genes was in Europe, followed by Asia, Africa, and the Americas. In general, the difference between two European populations separated by 1000 km is far less than in other world populations [47]. Geographical isolation together with various selection forces leads to an increase in F_{ST} values among human populations [48].

Locus-by-locus AMOVA revealed 11 SNPs with F_{ST} values greater than 10%. After calculating the MAFs of these 11 SNPs, we observed a clear clustering of SNPs in relation to

the studied populations: four SNPs were present in all world populations, but the most frequent in East Asian populations (rs2267447, rs1065852, rs2004511, and rs1081003), six SNPs were almost exclusively found in African populations (rs75203276, rs59421388, rs61736512, rs76327133, rs80262685, and rs28371706), while rs16947 was found in all world populations.

Four SNPs of the *CYP2D6* gene characteristic of East Asian populations occur in Vietnamese and three Chinese populations with a frequency in the range of 60–68%, and the Japanese population in the range of 36–39%. The finding that the Japanese population has frequencies that deviate from the rest of the group is not surprising since Japan regularly diverges due to their relative isolation throughout history. The rs1081003, a synonymous variant (Phe112Phe), showed the highest overall F_{ST} value of 0.37, and due to its substantially higher frequency in East Asian populations, it can be considered a typical East Asian variant. In addition, this SNP's minor A allele was also found to be the major allele in some South East Asian populations [49]. The Pharmacogene Variation (PharmVar) Consortium database defines rs1081003 as a suballele of numerous star alleles (*2, *4, *10, *36, *37, etc.).

SNP rs2004511, an intron variant with an overall F_{ST} value of 0.21, is also predominant in East Asian populations. In 2018, the association of a minor C allele with response to Tramadol was recorded in the ClinVar archive. The PharmVar defines that this SNP determines suballeles for a number of star alleles (*4, *10, *36, *39, etc.).

SNP rs1065852 in East Asian populations (except Japan) has frequencies of its A allele over 60%, making the G allele a minor allele in these populations. In addition to the four East Asian populations represented in the 1000 Genomes, its high frequency was also noticed in populations of Lisu [50] and Wa, both from Yunnan Province of China [51]. According to the PharmGKB, the A allele which causes a missense variant is associated with decreased clearance of alpha-hydroxymetoprolol in healthy individuals compared to the G allele. It is also associated with S-didesmethyl-citalopram plasma concentrations when treated with citalopram or escitalopram in people with depressive disorder. The GG genotype of rs1065852 is associated with a prolonged QTc interval when treating individuals with schizophrenia with iloperidone. Lee et al. (2016), analyzing the association between CYP polymorphisms and blood concentrations of hydroxychloroquine (HCQ) and its metabolite N-desethyl HCQ (DHCQ) in Korean patients with lupus, observed that patients with the GG genotype of allele *10 had the highest [DHCQ]/[HCQ] ratio, while patients with genotype AA had the lowest ratio [52]. López-García et al. (2017) found that this SNP, as it is included in the star allele *4, can affect the effectiveness of antiepileptic drugs [53]. Together with the SNP rs1081003, rs1065852 is in high LD in the populations of the Philippines, Thailand, Vietnam and Laos, and mutations in these key SNPs that define the star alleles *10 and *54 cause reduced CYP2D6 enzyme activity [49]. This SNP is part of the core of numerous star alleles (*10, *36, *37, etc.). The fourth SNP typical for East Asia is rs2267447, whose minor C allele causes change in the intron variant and is associated with response to Tramadol. According to the PharmVar, this SNP defines suballeles for numerous star alleles (*4, *10, *36, *39, etc.).

SNPs characteristic for the African group of populations were almost completely absent in other populations. rs28371706, whose minor A allele causes a missense variant, is the core SNP for defining star alleles *17, *40, *58, *64, *82, *141, and *154. The *CYP2D6*17* star allele occurs in at least 30% of Africans [54,55], and is associated with reduced enzyme activity-individuals carrying the *17 allele that are classified as intermediate metabolizers (IM). According to ClinVar, rs28371706 is associated with response to Tamoxifen and Deutetrabenazine.

The remaining five SNPs characteristic for the African populations have very similar MAF frequencies within each population. In PharmVar, SNPs rs80262685 (T > C, intron variant) and rs76327133 (G > A, intron variant) are associated with suballeles *2, *29, *146, *149, *155, *156, and *157, while the intron variant rs75203276 (C > T) is associated with suballeles *29, *155, *156, and *157. SNPs rs61736512 and rs59421388 (both C > T, missense variants) define the core of several star alleles; both define alleles *29, *70, *149, *155, *156, and *157, while rs61736512 also defines allele *107, and rs59421388 defines allele *109. Those

two variants are significantly associated with the decreased metabolism of debrisoquine, according to the PharmaGKB.

The distribution of rs16947 is the most intriguing. If we exclude the population of African ancestry from the Southwest USA where MAF was 42%, its MAF ranged from 50–65% in African populations, to 13–17% in East Asian populations. Muaymbo et al. (2022) found that this SNP had a significantly lower MAF among Africans in Southern Africa (12%) compared to their counterparts in West (65%) and East (56%) Africa [56]. The MAF of rs16947 in Southern Africans is lower than in any other world population, but is closest to the frequencies in East Asian populations. The rs16947 mutation (G > A), causing a missense variant which can result in decreased CYP2D6 enzyme activity, is one of the major mutations that distinguish star alleles *10A and *54 [49]. In the ClinVar archive, this SNP is associated with the Tamoxifen response, the Deitetrabenazine response, and the ultrarapid metabolism of Debrisoquine. The PharmVar defines that this SNP determines a number of core alleles.

The distribution of MAFs of the 11 SNP variants distinguishes African and East Asian populations from others. Haplotype-based PCA analyses showed separate clusters of African and East Asian populations, while South Asian, European and American populations were much less separated. Separation of African populations follows their genetic history and is visible in different genetic studies as well [57,58]. The clustering of East Asian populations is also evident from studies of other pharmacogenes. Li et al. demonstrated that ADME genes exhibit distinct patterns of population differentiation in a global and regional context. While some genes are conserved (e.g., *SLC04C1* and *NAT1*), others (e.g., *CYP3A5*) exhibit high levels of world population differentiation. On the other hand, the global diversity of some genes primarily reflects differentiation within a particular geographical region, such as Africa, Europe, or East Asia [59]. The genomic diversity of modern populations reflects former demographic and evolutionary changes. In isolated populations with minimal gene exchange, genetic distinctiveness is especially evident (e.g., Jewish populations, Saami, Roma, Basque), which can also be seen in pharmacogene research [60–66].

Population variation has been studied through patterns of haplotype blocks. African populations have the most diverse block pattern, while it is the most homogeneous in Europe, followed by the Americas, and South and East Asia. The investigated haplotype blocks, based on confidence intervals [39], encompass the region of SNPs with strong LD as a consequence of lack of recombination. The largest number of blocks present in African populations is consistent with their genetic history. The African population has had a relatively large effective population size over a long period of time, allowing recombination events to leave their mark on the haplotype block structure. The significant correlation between SNPs defining the ends of haplotype blocks further supports their informativeness in the African population.

Haplotypes containing functional/associated variants are more likely to determine clinical drug metabolism phenotypes than a single independent SNP [67]. Among the present investigated haplotypes translated into star alleles nomenclature, haplotype 1 defines *1 allele, which is a normal metabolizer and is the most common haplotype in all populations outside of East Asia, where it is the second most common. Haplotype 2 defines *2 allele and is the second most abundant in European, South Asian, and American populations. Haplotype 3 defines *10 allele and is very characteristic of East Asian populations. Haplotype 4 defines *4 allele and appears in European and American populations. Haplotype 5 defines *41 allele and occurs in South Asian and European populations. Haplotypes 6 (*29), 7 (*1), and 8 (*17) occur only in African populations.

Sistonen and colleagues demonstrated different distributions of the CYP2D6 slow (i.e., *9, *10, *17, *29, *45–46) and null-function (i.e., *4, *5, *6) alleles on different continents, probably caused by demographic events [68]. Decreased function CYP2D6 enzymes are characterized by substrate-dependent catalytic properties (gene variants *10, *17, and *29) and enzyme inhibitor affinities (*10, *17), which contribute to a broader spectrum

of metabolic responses [69–71]. Individuals defined as poor (slow) metabolizers may metabolize certain classes of chemical compounds better (or worse), which should not necessarily be unfavorable. For example, if toxic compounds were activated through CYP2D6-mediated metabolism, slow metabolizers could be at reduced risk of adverse effects [31]. Many widely used therapeutic medications, drugs of abuse, exogenous chemicals such as alkaloids, herbicides, and some endogenous molecules such as progesterone, estrogen, and many other substances are substrates for the human CYP2D6 enzyme [72].

5. Conclusions

African populations showed the highest variability of the *CYP2D6* haplotypes, which is consistent with all known studies of genetic variability in humans. However, the greatest differentiation is found among East Asian populations due to extreme homogeneity of the Japanese population and the specific distribution of haplotypes among the Chinese population. The greatest continental homogeneity is found in Europe, followed by South Asia and the Americas.

Locus-by-locus analyses revealed 11 SNP loci affecting inter-population differentiation, six of which are specific to African, four to East Asian populations, while one is present globally. Five of these SNPs (rs2004511, rs1065852, rs2267447, rs28371706, and rs16947) contribute to the known pharmacogenomic effects on clinical outcomes of drugs metabolized by CYP2D6.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pharmaceutics14112481/s1, Figure S1: Haplotype blocks for 26 world populations; Table S1: Allele frequencies from 26 world populations available in the 1000 Genomes database.

Author Contributions: Conceptualization, M.P.S. and M.Z.P.; methodology, A.S.M. and M.P.S.; validation, all authors; formal analysis, A.S.M. and M.Z.P.; investigation, A.S.M., M.P.S. and M.Z.P.; data curation, T.Š.-J., M.Z.P. and Ž.T.; writing—original draft preparation, A.S.M. and M.Z.P.; writing—review and editing, all authors; visualization, Ž.C. and M.Š.; supervision, M.P.S.; funding acquisition, M.P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Croatian Science Foundation (IP-2014-09-4454 and DOK-2018-01-4817 to M.P.S.).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Scientific Committee and the Ethics Committee of the Institute for Anthropological Research (RN 1.14-1611/14).

Data Availability Statement: All data analyzed in this study are available at http://roma.inantro.hr/en/, (accessed on 5 October 2022). In the case of using this database for further analyses, please cite this publication. If further clarification is required, contact the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Larrey, D.; Distlerath, L.M.; Dannan, G.A.; Wilkinson, G.R.; Guengerich, F.P. Purification and characterization of the rat liver microsomal cytochrome P-450 involved in the 4-hydroxylation of debrisoquine, a prototype for genetic variation in oxidative drug metabolism. *Biochemistry* **1984**, *23*, 2787–2795. [CrossRef] [PubMed]
2. Eichelbaum, M.; Baur, M.; Dengler, H.; Osikowska-Evers, B.; Tieves, G.; Zekorn, C.; Rittner, C. Chromosomal assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. *Br. J. Clin. Pharmacol.* **1987**, *23*, 455–458. [CrossRef] [PubMed]
3. Kimura, S.; Umeno, M.; Skoda, R.C.; A Meyer, U.; Gonzalez, F.J. The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am. J. Hum. Genet.* **1989**, *45*, 889–904. [PubMed]
4. Daly, A.K.; Brockmoller, J.; Broly, F.; Eichelbaum, M.; E Evans, W.; Gonzalez, F.J.; Huang, J.D.; Idle, J.R.; Ingelman-Sundberg, M.; Ishizaki, T.; et al. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics* **1996**, *6*, 193–201. [CrossRef] [PubMed]
5. Gresham, D.; Morar, B.; Underhill, P.A.; Passarino, G.; Lin, A.A.; Wise, C.; Angelicheva, D.; Calafell, F.; Oefner, P.J.; Shen, P.; et al. Origins and Divergence of the Roma (Gypsies). *Am. J. Hum. Genet.* **2001**, *69*, 1314–1331. [CrossRef] [PubMed]

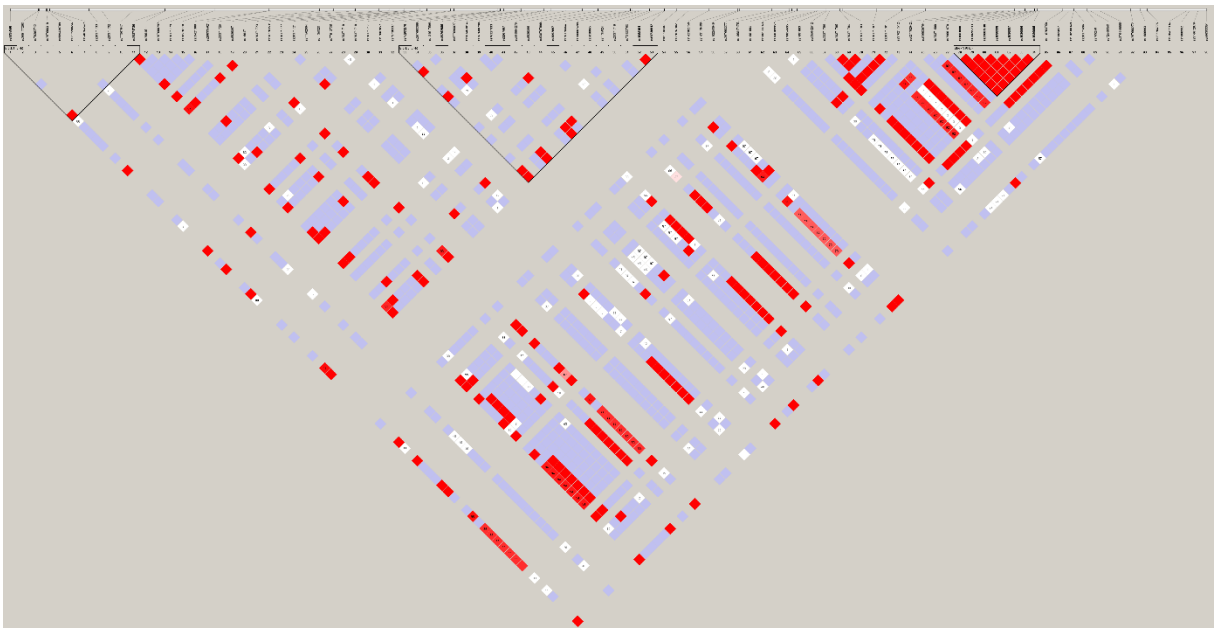
6. Taylor, C.; Crosby, I.; Yip, V.; Maguire, P.; Pirmohamed, M.; Turner, R.M. A Review of the Important Role of *CYP2D6* in Pharmacogenomics. *Genes* **2020**, *11*, 1295. [CrossRef] [PubMed]
7. Caudle, K.E.; Sangkuhl, K.; Whirl-Carrillo, M.; Swen, J.J.; Haidar, C.E.; Klein, T.E.; Gammal, R.S.; Relling, M.V.; Scott, S.A.; Hertz, D.L.; et al. Standardizing CYP 2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clin. Transl. Sci.* **2020**, *13*, 116–124. [CrossRef]
8. The Human Cythchrome P450 (CYP) Allele Nomenclature Database. Available online: http://cypalleles.ki.se/ (accessed on 15 April 2022).
9. PHARMGKB. Stanford University. Website. Available online: https://www.pharmgkb.org (accessed on 13 December 2021).
10. Hicks, J.K.; Swen, J.J.; Gaedigk, A. Challenges in *CYP2D6* Phenotype Assignment from Genotype Data: A Critical Assessment and Call for Standardization. *Curr. Drug Metab.* **2014**, *15*, 218–232. [CrossRef]
11. Zanger, U.M.; Schwab, M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* **2013**, *138*, 103–141. [CrossRef]
12. Williams, I.S.; Gatchie, L.; Bharate, S.B.; Chaudhuri, B. Biotransformation, Using Recombinant CYP450-Expressing Baker's Yeast Cells, Identifies a Novel CYP2D6.10A122V Variant Which Is a Superior Metabolizer of Codeine to Morphine Than the Wild-Type Enzyme. *ACS Omega* **2018**, *3*, 8903–8912. [CrossRef] [PubMed]
13. Ingelman-Sundberg, M.; Sim, S.C.; Gomez, A.; Rodriguez-Antona, C. Influence of cytochrome P450 polymor-phisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.* **2007**, *116*, 496–526. [CrossRef] [PubMed]
14. Zanger, U.M.; Turpeinen, M.; Klein, K.; Schwab, M. Functional pharmacogenetics/genomics of human cyto-chromes P450 involved in drug biotransformation. *Anal. Bioanal. Chem.* **2008**, *392*, 1093–1108. [CrossRef] [PubMed]
15. Fleeman, N.; Dundar, Y.; Dickson, R.; Jorgensen, A.; Pushpakom, S.; McLeod, C.; Pirmohamed, M.; Walley, T. Cytochrome P450 testing for prescribing antipsychotics in adults with schizophrenia: Systematic review and meta-analyses. *Pharm. J.* **2010**, *11*, 1–14. [CrossRef] [PubMed]
16. Stingl, J.C.; Brockmöller, J.; Viviani, R. Genetic variability of drug-metabolizing enzymes: The dual impact on psychiatric therapy and regulation of brain function. *Mol. Psychiatry* **2012**, *18*, 273–287. [CrossRef] [PubMed]
17. Gaedigk, A. Complexities of *CYP2D6* gene analysis and interpretation. *Int. Rev. Psychiatry* **2013**, *25*, 534–553. [CrossRef] [PubMed]
18. Hicks, J.K.; Bishop, J.R.; Sangkuhl, K.; Müller, D.J.; Ji, Y.; Leckband, S.G.; Leeder, J.S.; Graham, R.L.; Chiulli, D.L.; Llerena, A.; et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for *CYP2D6* and *CYP2C19* Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clin. Pharmacol. Ther.* **2015**, *98*, 127–134. [CrossRef]
19. Beoris, M.; Wilson, J.A.; Garces, J.A.; Lukowiak, A.A. CYP2D6 copy number distribution in the US population. *Pharm. Genom.* **2016**, *26*, 96–99. [CrossRef]
20. Christensen, P.M.; Gotzsche, P.C.; Brosen, K. The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of Parkinson's disease: A meta-analysis. *Pharmacogenetics* **1998**, *8*, 473–479. [CrossRef]
21. Lu, Y.; Peng, Q.; Zeng, Z.; Wang, J.; Deng, Y.; Xie, L.; Mo, C.; Zeng, J.; Qin, X.; Li, S. CYP2D6 phenotypes and Parkinson's disease risk: A meta-analysis. *J. Neurol. Sci.* **2014**, *336*, 161–168. [CrossRef]
22. Mishra, A.K.; Singh, M.P. Cytochrome P450 2D6 and Parkinson's Disease: Polymorphism, Metabolic Role, Risk and Protection. *Neurochem. Res.* **2017**, *42*, 3353–3361. [CrossRef]
23. Patsopoulos, N.; Ntzani, E.E.; Zintzaras, E.; Ioannidis, J.P. CYP2D6 polymorphisms and the risk of tardive dyskinesia in schizophrenia: A meta-analysis. *Pharm. Genom.* **2005**, *15*, 151–158. [CrossRef] [PubMed]
24. Scordo, M.G.; Dahl, M.-L.; Spina, E.; Cordici, F.; Arena, M.G. No association between CYP2D6 polymorphism and Alzheimer's disease in an Italian population. *Pharmacol. Res.* **2006**, *53*, 162–165. [CrossRef] [PubMed]
25. Ma, S.L.; Tang, N.L.S.; Wat, K.H.Y.; Tang, J.H.Y.; Lau, K.H.; Law, C.B.; Chiu, J.; Tam, C.C.W.; Poon, T.K.; Lin, K.L.; et al. Effect of CYP2D6 and CYP3A4 Genotypes on the Efficacy of Cholinesterase Inhibitors in Southern Chinese Patients With Alzheimer's Disease. *Am. J. Alzheimer's Dis. Other Dement.* **2019**, *34*, 302–307. [CrossRef] [PubMed]
26. Agundez, J.A. Cytochrome P450 Gene Polymorphism and Cancer. *Curr. Drug Metab.* **2004**, *5*, 211–224. [CrossRef]
27. Rodriguez-Antona, C.; Gomez, A.; Karlgren, M.; Sim, S.C.; Ingelman-Sundberg, M. Molecular genetics and epigenetics of the cytochrome P450 gene family and its relevance for cancer risk and treatment. *Qual. Life Res.* **2009**, *127*, 1–17. [CrossRef]
28. Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. *Pharm. J.* **2004**, *5*, 6–13. [CrossRef]
29. Aklillu, E.; Herrlin, K.; Gustafsson, L.L.; Bertilsson, L.; Ingelman-Sundberg, M. Evidence for environmental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. *Pharmacogenetics* **2002**, *12*, 375–383. [CrossRef]
30. Podgorná, E.; Diallo, I.; Vangenot, C.; Sanchez-Mazas, A.; Sabbagh, A.; Černý, V.; Poloni, E.S. Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. *BMC Evol. Biol.* **2015**, *15*, 263. [CrossRef]
31. Fuselli, S.; de Filippo, C.; Mona, S.; Sistonen, J.; Fariselli, P.; Destro-Bisol, G.; Barbujani, G.; Bertorelle, G.; Sajantila, A. Evolution of detoxifying systems: The role of environment and population history in shaping genetic diversity at human CYP2D6 locus. *Pharm. Genom.* **2010**, *20*, 485–499. [CrossRef]

32. Thomas, J.H. Rapid Birth–Death Evolution Specific to Xenobiotic Cytochrome P450 Genes in Vertebrates. *PLoS Genet.* **2007**, *3*, e67. [CrossRef]
33. Delser, P.M.; Fuselli, S. Human loci involved in drug biotransformation: Worldwide genetic variation, population structure, and pharmacogenetic implications. *Qual. Life Res.* **2013**, *132*, 563–577. [CrossRef]
34. Fuselli, S. Beyond drugs: The evolution of genes involved in human response to medications. *Proc. R. Soc. B Boil. Sci.* **2019**, *286*, 20191716. [CrossRef] [PubMed]
35. Cunningham, F.; E Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [CrossRef]
36. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [CrossRef] [PubMed]
37. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
38. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [CrossRef]
39. Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.M.; Roy, J.; Blumenstiel, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; et al. The Structure of Haplotype Blocks in the Human Genome. *Science* **2002**, *296*, 2225–2229. [CrossRef]
40. Stephens, M.; Donnelly, P. A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *Am. J. Hum. Genet.* **2003**, *73*, 1162–1169. [CrossRef]
41. Stephens, M.; Smith, N.J.; Donnelly, P. A New Statistical Method for Haplotype Reconstruction from Population Data. *Am. J. Hum. Genet.* **2001**, *68*, 978–989. [CrossRef]
42. Excoffier, L.; Lischer, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [CrossRef]
43. van der Weide, J.; Steijns, L.S.W. Cytochrome P450 Enzyme System: Genetic Polymorphisms and Impact on Clinical Pharmacology. *Ann. Clin. Biochem. Int. J. Lab. Med.* **1999**, *36*, 722–729. [CrossRef] [PubMed]
44. Rodrigues, J.C.G.; Fernandes, M.R.; Ribeiro-Dos-Santos, A.M.; de Araújo, G.S.; de Souza, S.J.; Guerreiro, J.F.; Ribeiro-Dos-Santos, Â.; de Assumpção, P.P.; dos Santos, N.P.C.; Santos, S. Pharmacogenomic Profile of Amazonian Amerindians. *J. Pers. Med.* **2022**, *12*, 952. [CrossRef] [PubMed]
45. Idda, M.L.; Zoledziewska, M.; Urru, S.A.M.; McInnes, G.; Bilotta, A.; Nuvoli, V.; Lodde, V.; Orrù, S.; Schlessinger, D.; Cucca, F.; et al. Genetic Variation among Pharmacogenes in the Sardinian Population. *Int. J. Mol. Sci.* **2022**, *23*, 10058. [CrossRef] [PubMed]
46. Wang, W.Y.; Twesigomwe, D.; Nofziger, C.; Turner, A.J.; Helmecke, L.-S.; Broeckel, U.; Derezinski, A.D.; Hazelhurst, S.; Gaedigk, A. Characterization of Novel *CYP2D6* Alleles across Sub-Saharan African Populations. *J. Pers. Med.* **2022**, *12*, 1575. [CrossRef] [PubMed]
47. Jay, F.; Sjödin, P.; Jakobsson, M.; Blum, M.G. Anisotropic Isolation by Distance: The Main Orientations of Human Genetic Differentiation. *Mol. Biol. Evol.* **2013**, *30*, 513–525. [CrossRef] [PubMed]
48. Bamshad, M.; Wooding, S.P. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **2003**, *4*, 99–110. [CrossRef] [PubMed]
49. Runcharoen, C.; Fukunaga, K.; Sensorn, I.; Iemwimangsa, N.; Klumsathian, S.; Tong, H.; Vo, N.S.; Le, L.; Hlaing, T.M.; Thant, M.; et al. Prevalence of pharmacogenomic variants in 100 pharmacogenes among Southeast Asian populations under the collaboration of the Southeast Asian Pharmacogenomics Research Network (SEAPharm). *Hum. Genome Var.* **2021**, *8*, 1–6. [CrossRef]
50. Zhang, C.; Jiang, X.; Chen, W.; Li, Q.; Yun, F.; Yang, X.; Dai, R.; Cheng, Y. Population genetic difference of pharmacogenomic VIP gene variants in the Lisu population from Yunnan Province. *Medicine* **2018**, *97*, e13674. [CrossRef]
51. Li, D.; Peng, L.; Xing, S.; He, C.; Jin, T. Genetic analysis of pharmacogenomic VIP variants in the Wa population from Yunnan Province of China. *BMC Genom. Data* **2021**, *22*, 1–20. [CrossRef]
52. Lee, J.Y.; Vinayagamoorthy, N.; Han, K.; Kwok, S.K.; Ju, J.H.; Park, K.S.; Jung, S.-H.; Park, S.-W.; Chung, Y.-J. Association of Polymorphisms of Cytochrome P450 2D6 With Blood Hydroxychloroquine Levels in Patients with Systemic Lupus Erythematosus. *Arthritis Rheumatol.* **2016**, *68*, 184–190. [CrossRef]
53. López-García, M.A.; Feria-Romero, I.A.; Serrano, H.; Rayo-Mares, D.; Fagiolino, P.; Vázquez, M.; Escamilla-Núñez, C.; Grijalva, I.; Escalante-Santiago, D.; Orozco-Suarez, S. Influence of genetic variants of CYP2D6, CYP2C9, CYP2C19 and CYP3A4 on antiepileptic drug metabolism in pediatric patients with refractory epilepsy. *Pharmacol. Rep.* **2017**, *69*, 504–511. [CrossRef] [PubMed]
54. Lymperopoulos, A.; McCrink, K.A.; Brill, A. Impact of CYP2D6 Genetic Variation on the Response of the Cardiovascular Patient to Carvedilol and Metoprolol. *Curr. Drug Metab.* **2015**, *17*, 30–36. [CrossRef] [PubMed]
55. Masimirembwa, C.; Persson, I.; Bertilsson, L.; Hasler, J.; Ingelman-Sundberg, M. A novel mutant variant of the *CYP2D6* gene (CYP2D617) common in a black African population: Association with diminished debrisoquine hydroxylase activity. *Br. J. Clin. Pharmacol.* **1996**, *42*, 713–719. [CrossRef]
56. Muyambo, S.; Ndadza, A.; Soko, N.D.; Kruger, B.; Kadzirange, G.; Chimusa, E.; Masimirembwa, C.M.; Ntsekhe, M.; Nhachi, C.F.; Dandara, C. Warfarin Pharmacogenomics for Precision Medicine in Real-Life Clinical Practice in Southern Africa: Harnessing 73 Variants in 29 Pharmacogenes. *OMICS: A J. Integr. Biol.* **2022**, *26*, 35–50. [CrossRef] [PubMed]

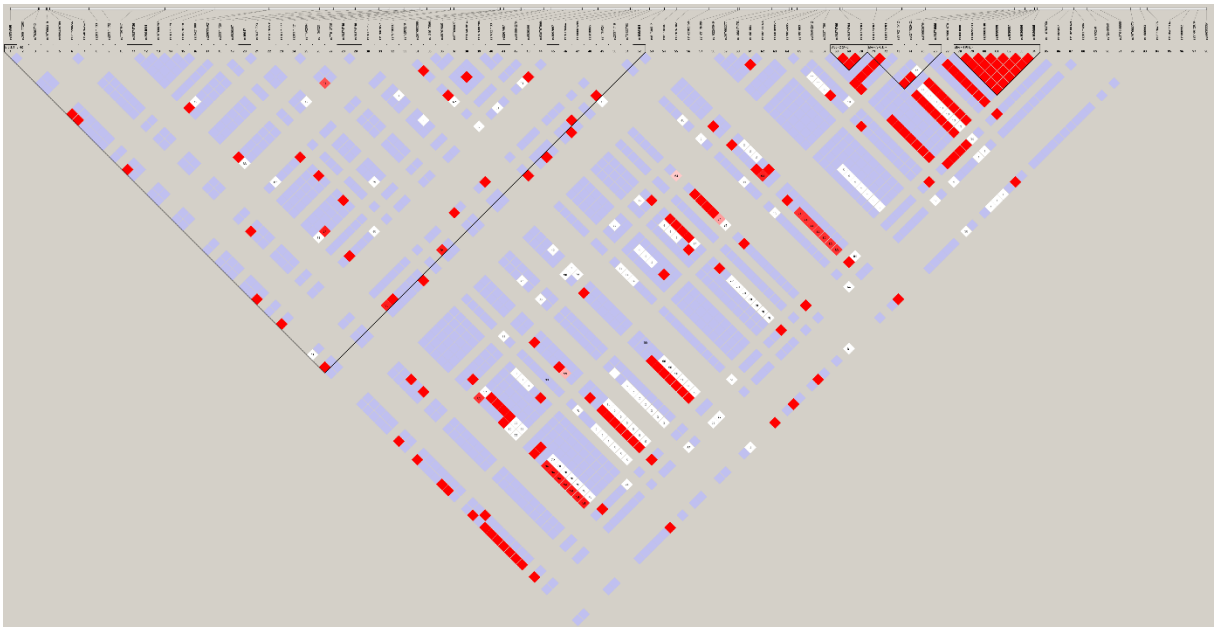
57. Prohaska, A.; Racimo, F.; Schork, A.J.; Sikora, M.; Stern, A.J.; Ilardo, M.; Allentoft, M.E.; Folkersen, L.; Buil, A.; Moreno-Mayar, J.V.; et al. Human Disease Variation in the Light of Population Genomics. *Cell* **2019**, *177*, 115–131. [CrossRef] [PubMed]
58. Sanchez-Mazas, A.; Poloni, E.S.; Genetic Diversity in Africa. Encyclopedia of Life Sciences 2008. Available online: https://www.els.net/ (accessed on 15 April 2022).
59. Li, J.; Zhang, L.; Zhou, H.; Stoneking, M.; Tang, K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* **2011**, *20*, 528–540. [CrossRef] [PubMed]
60. Marković, A.S.; Petranović, M.Z.; Škobalj, M.; Poloni, E.S.; Oberški, L.P.; Škarić-Jurić, T.; Salihović, M.P. From dietary adaptation in the past to drug metabolism of today: An example of NAT genes in the Croatian Roma. *Am. J. Biol. Anthr.* **2022**, *178*, 140–153. [CrossRef]
61. Marković, A.S.; Petranović, M.Z.; Tomas, Z.; Puljko, B.; Šetinc, M.; Škarić-Jurić, T.; Salihović, M.P. Untangling SNP Variations within *CYP2D6* Gene in Croatian Roma. *J. Pers. Med.* **2022**, *12*, 374. [CrossRef]
62. Petranovic, M.Z.; Tomas, Z.; Skaric-Juric, T.; Narancic, N.S.; Janicijevic, B.; Markovic, A.S.; Salihovic, M.P. The variability of multi-drug resistance ABCB1 gene in the Roma population from Croatia. *Mol. Exp. Biol. Med.* **2019**, *2*, 10–18. [CrossRef]
63. Dlouhá, L.; Adámková, V.; Šedová, L.; Olišarová, V.; Hubáček, J.A.; Tóthová, V. Five genetic polymorphisms of cytochrome P450 enzymes in the Czech non-Roma and Czech Roma population samples. *Drug Metab. Pers. Ther.* **2020**, *35*, 20200103. [CrossRef]
64. Weber, A.; Szalai, R.; Sipeky, C.; Magyari, L.; Melegh, M.; Jaromi, L.; Matyas, P.; Duga, B.; Kovesdi, E.; Hadzsiev, K.; et al. Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples. *Pharmacol. Rep.* **2015**, *67*, 460–464. [CrossRef] [PubMed]
65. Moyà, G.; Dorado, P.; Ferreiro, V.; Naranjo, M.E.G.; Peñas-Lledó, E.M.; Llerena, A. High frequency of CYP2D6 ultrarapid metabolizer genotypes in an Ashkenazi Jewish population from Argentina. *Pharm. J.* **2016**, *17*, 378–381. [CrossRef]
66. Wen, Y.F.; Gaedigk, A.; Boone, E.C.; Wang, W.Y.; Straka, R.J. The Identification of Novel CYP2D6 Variants in US Hmong: Results from Genome Sequencing and Clinical Genotyping. *Front. Pharmacol.* **2022**, *13*, 867331. [CrossRef] [PubMed]
67. Li, J.; Lou, H.; Yang, X.; Lu, D.; Li, S.; Jin, L.; Pan, X.; Yang, W.; Song, M.; Mamatyusupu, D.; et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *J. Med Genet.* **2014**, *51*, 614–622. [CrossRef] [PubMed]
68. Sistonen, J.; Sajantila, A.; Lao, O.; Corander, J.; Barbujani, G.; Fuselli, S. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharm. Genom.* **2007**, *17*, 93–101. [CrossRef]
69. Wennerholm, A.; Dandara, C.; Sayi, J.; Svensson, J.; Abdi, Y.A.; Ingelman-Sundberg, M.; Bertilsson, L.; Hasler, J.; Gustafsson, L.L. The African-specific CYP2D6*17 allele encodes an enzyme with changed substrate specificity. *Clin. Pharmacol. Ther.* **2002**, *71*, 77–88. [CrossRef]
70. Bogni, A.; Monshouwer, M.; Moscone, A.; Hidestrand, M.; Ingelman-Sundberg, M.; Hartung, T.; Coecke, S. Substrate specific metabolism by polymorphic cytochrome P450 2D6 alleles. *Toxicol. Vitr.* **2005**, *19*, 621–629. [CrossRef]
71. Shen, H.; He, M.M.; Liu, H.; Wrighton, S.A.; Wang, L.; Guo, B.; Li, C. Comparative Metabolic Capabilities and Inhibitory Profiles of CYP2D6.1, CYP2D6.10, and CYP2D6. *Drug Metab. Dispos.* **2007**, *35*, 1292–1300. [CrossRef]
72. Wang, B.; Yang, L.-P.; Zhang, X.-Z.; Huang, S.-Q.; Bartlam, M.; Zhou, S.-F. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. *Drug Metab. Rev.* **2009**, *41*, 573–643. [CrossRef]

Figure S1: Haplotype blocks for 26 world populations

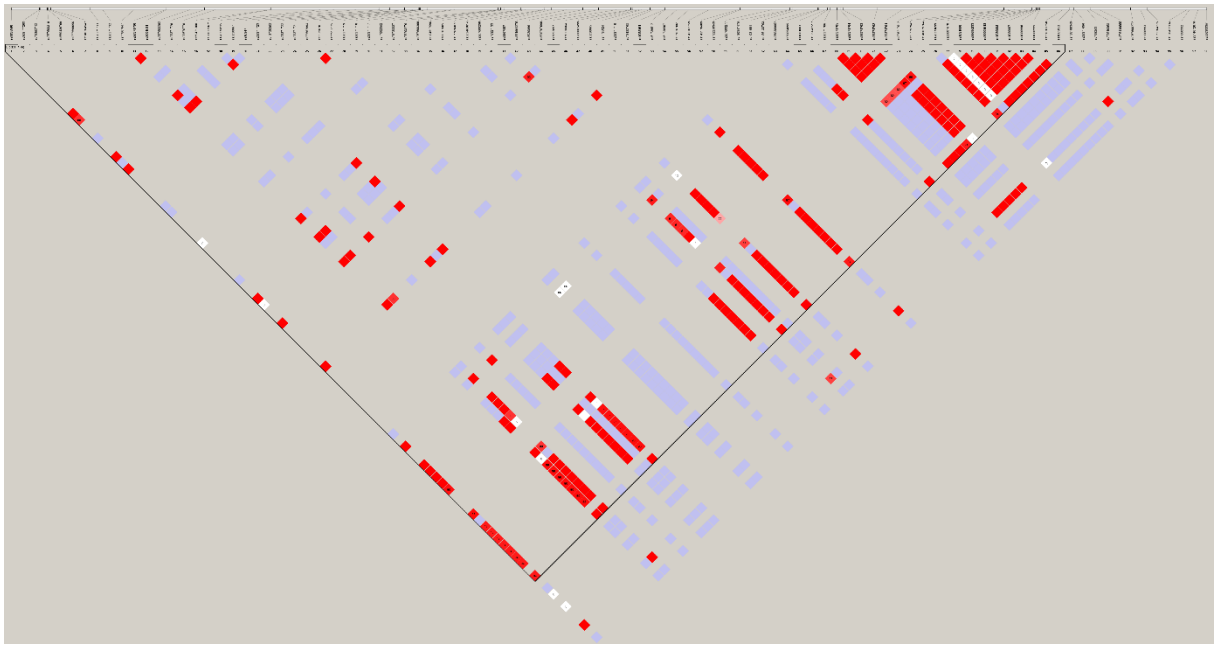
ACB



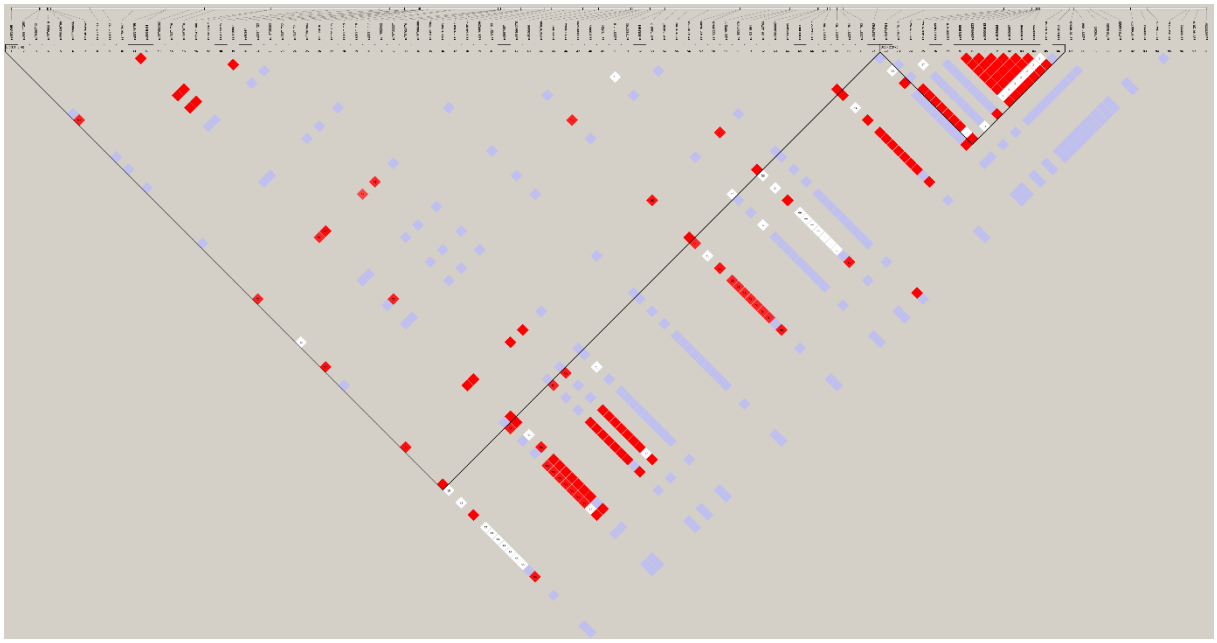
ASW



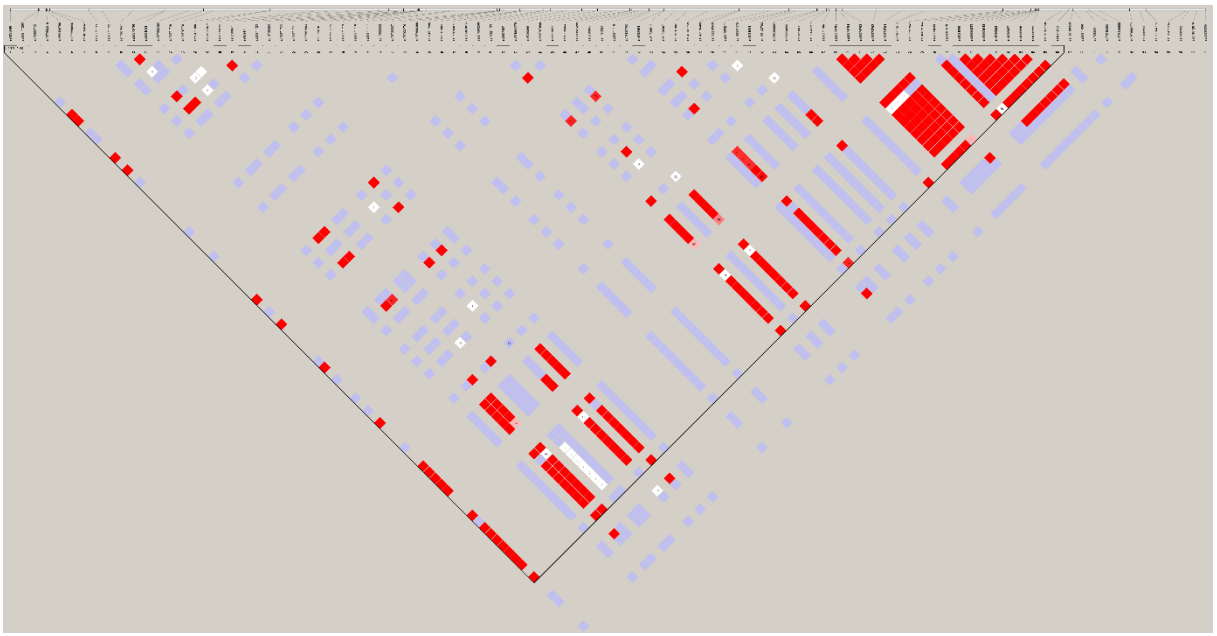
BEB



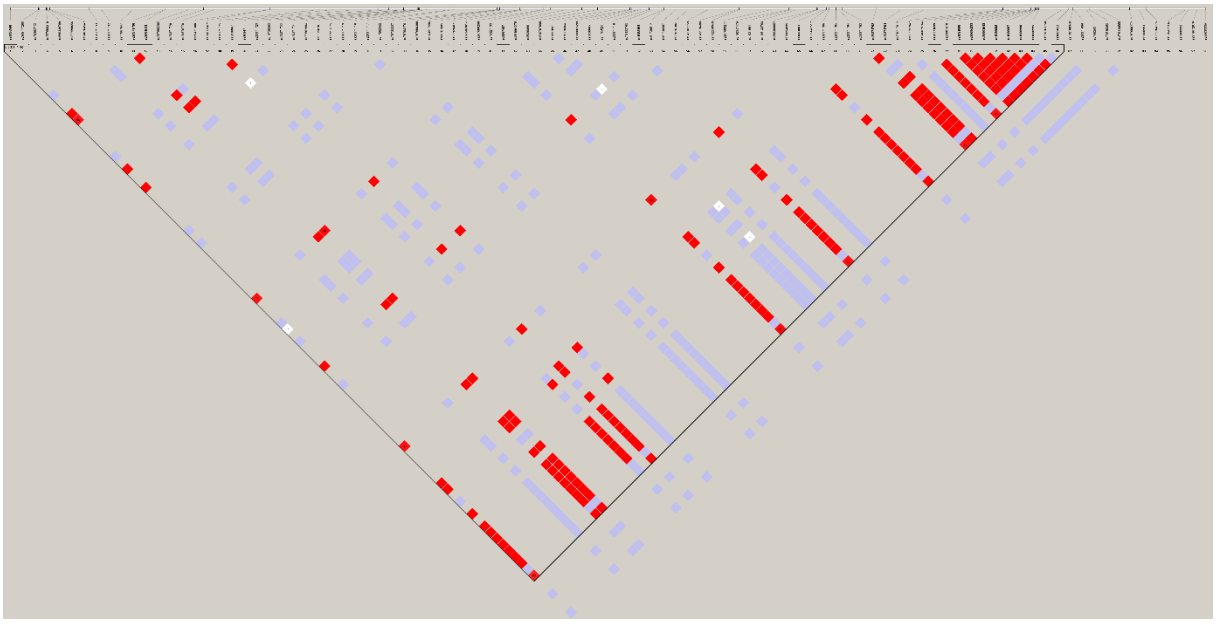
CDX



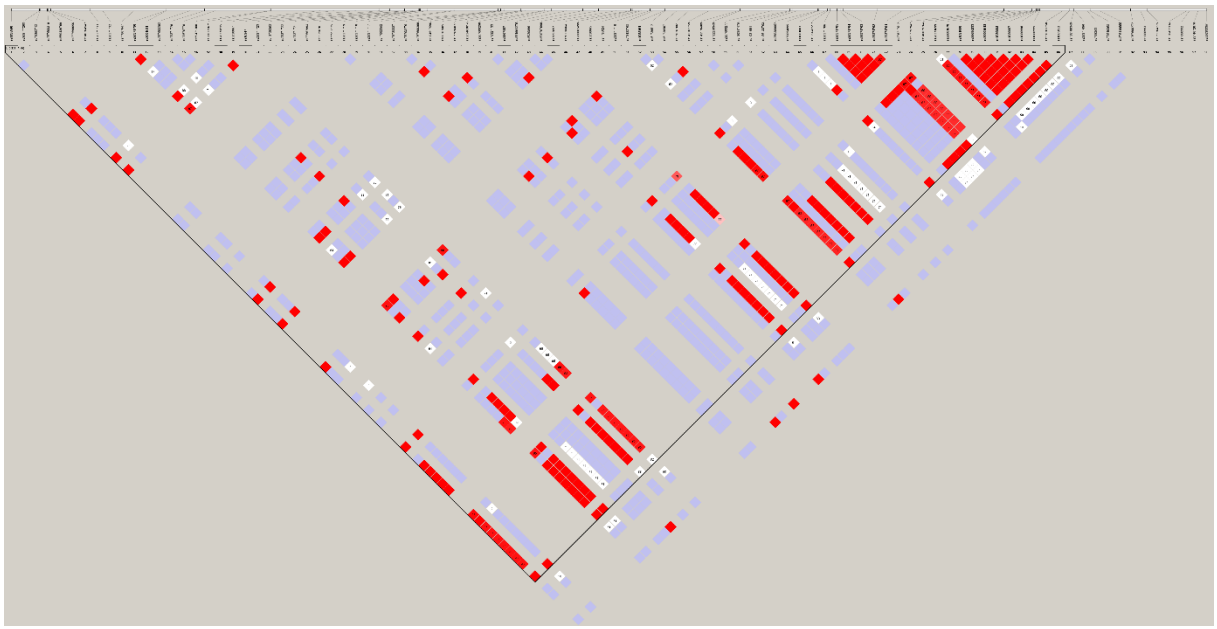
CEU



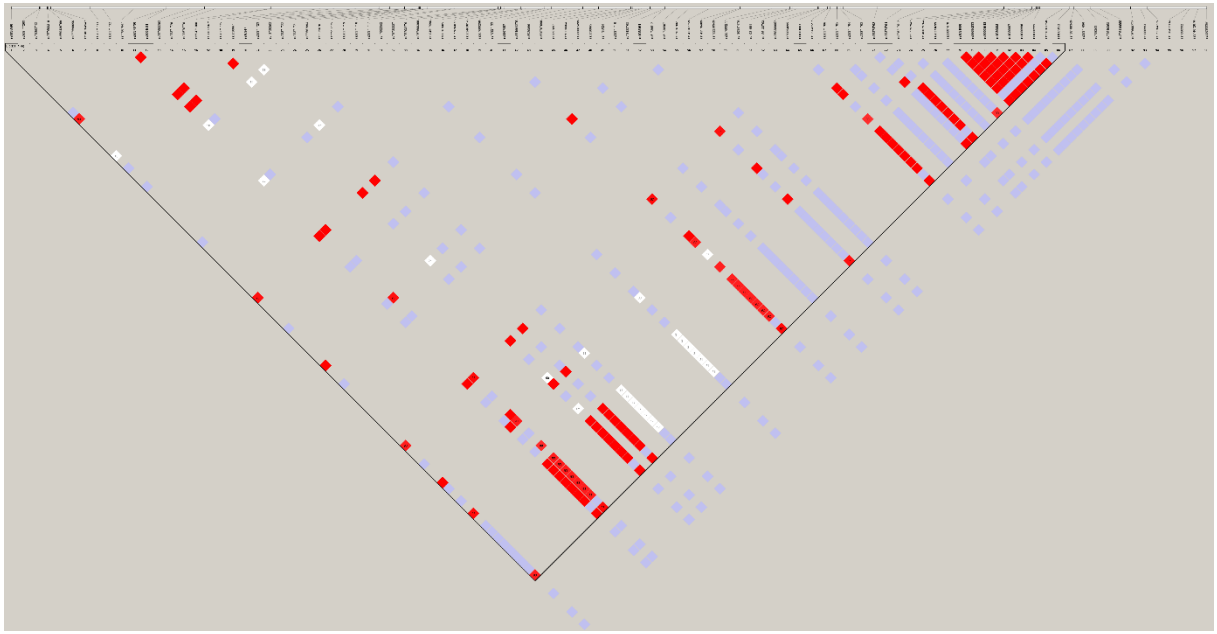
CHB



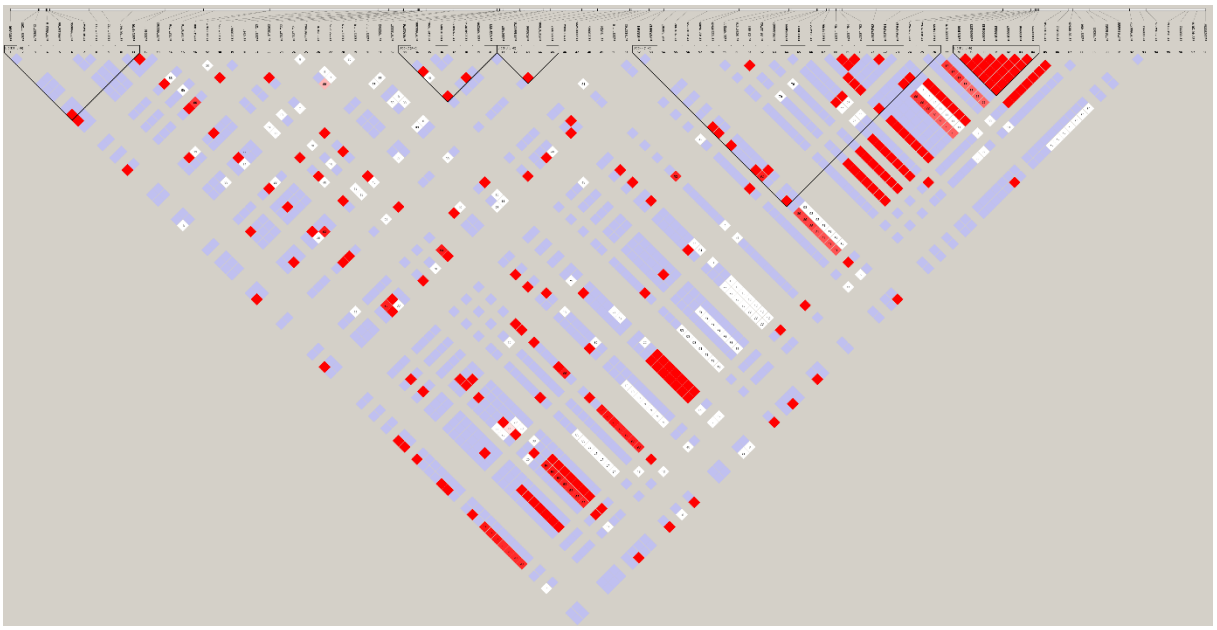
CLM



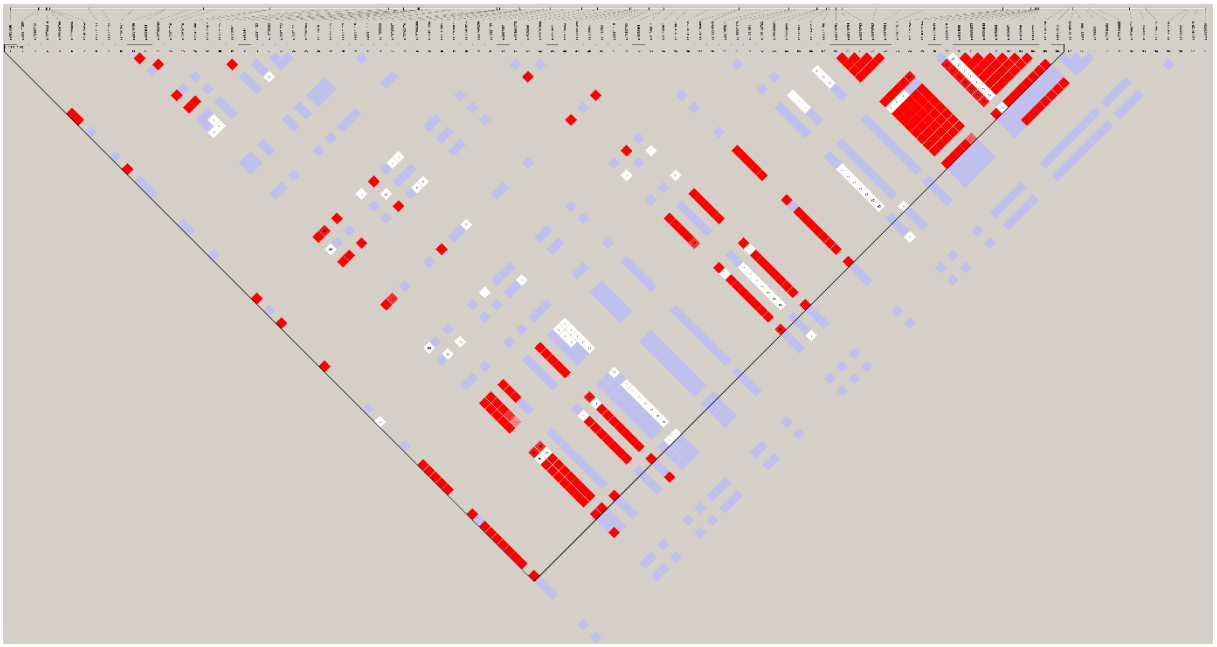
CHS



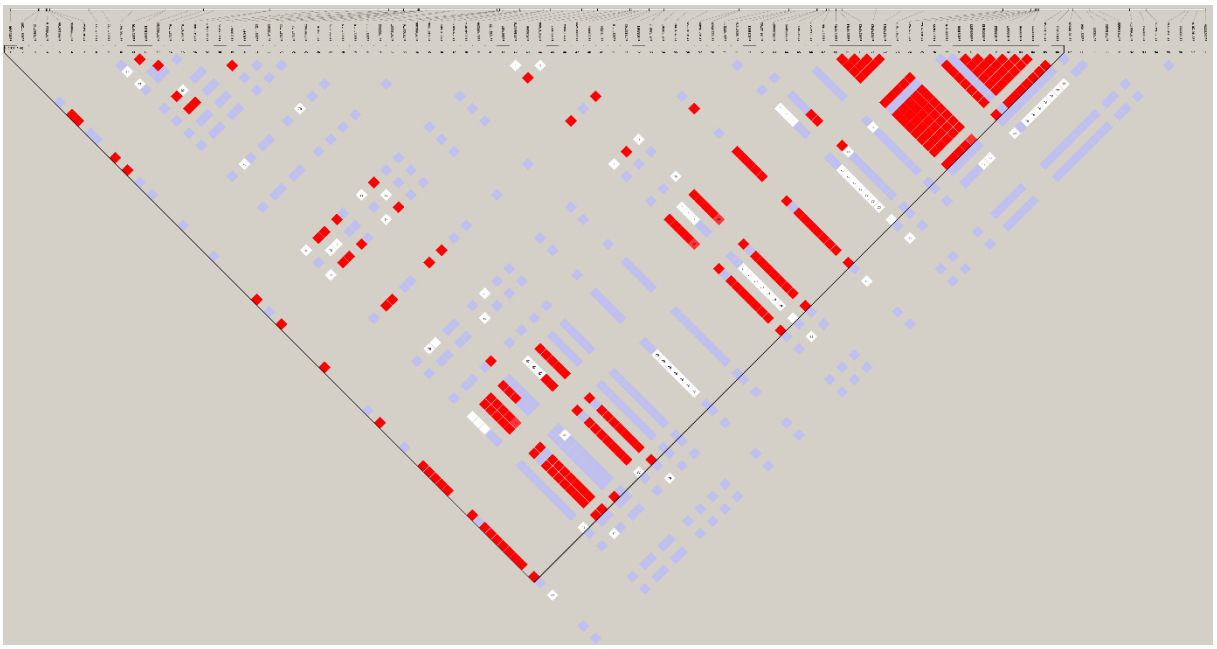
ESN



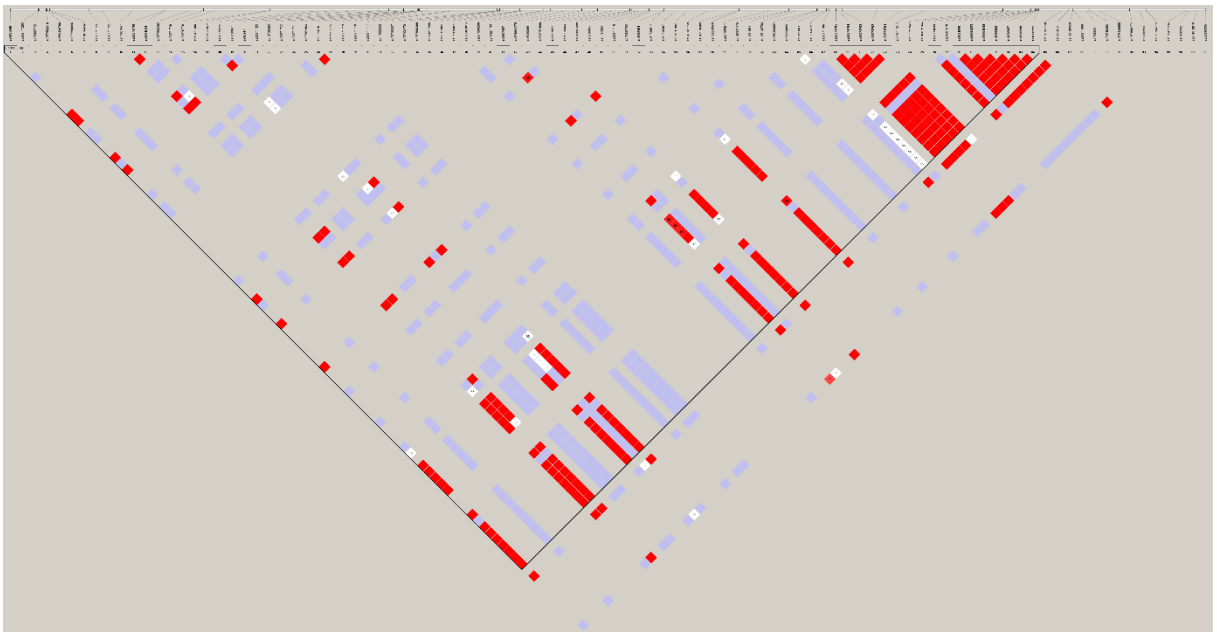
FIN



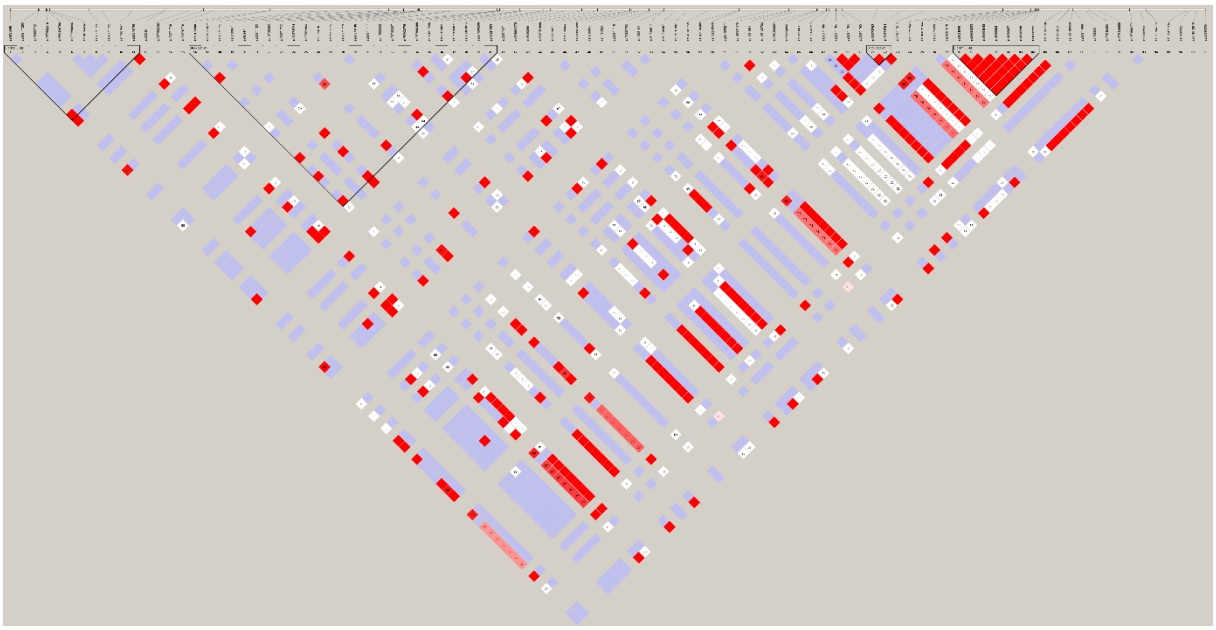
GBR



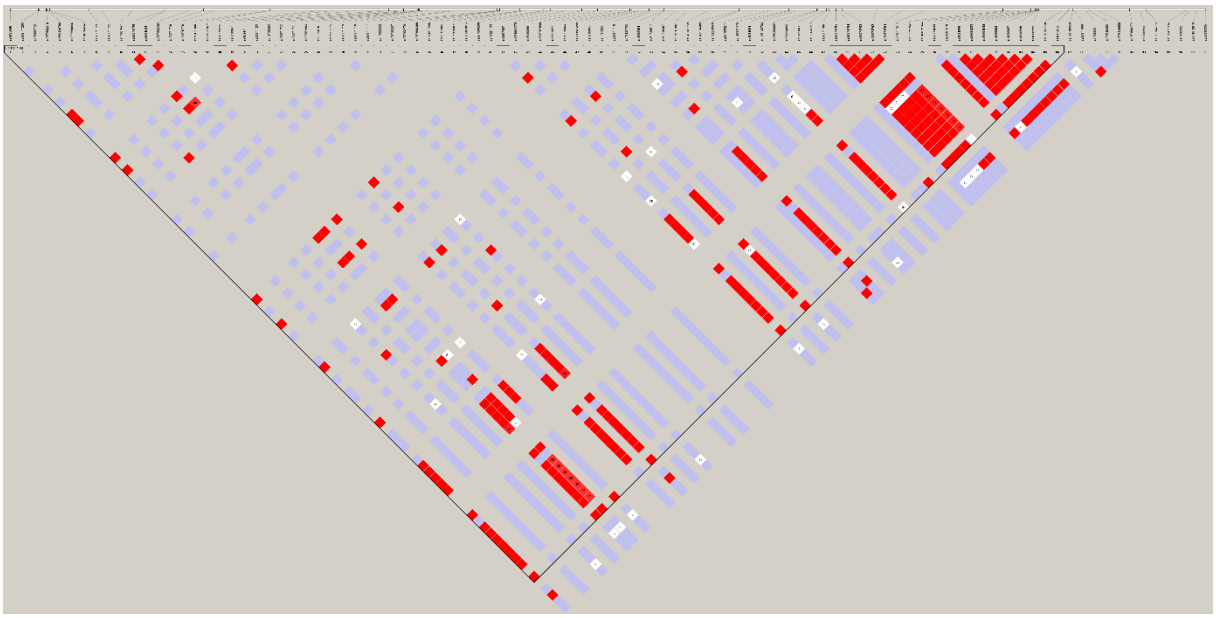
GIH



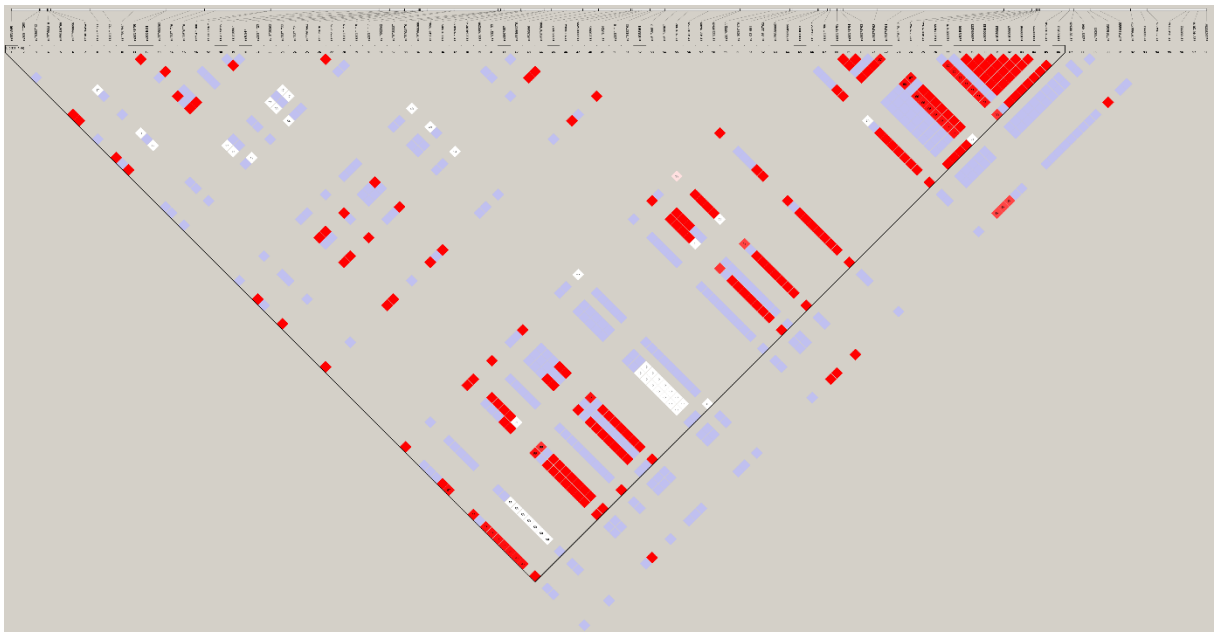
GWD



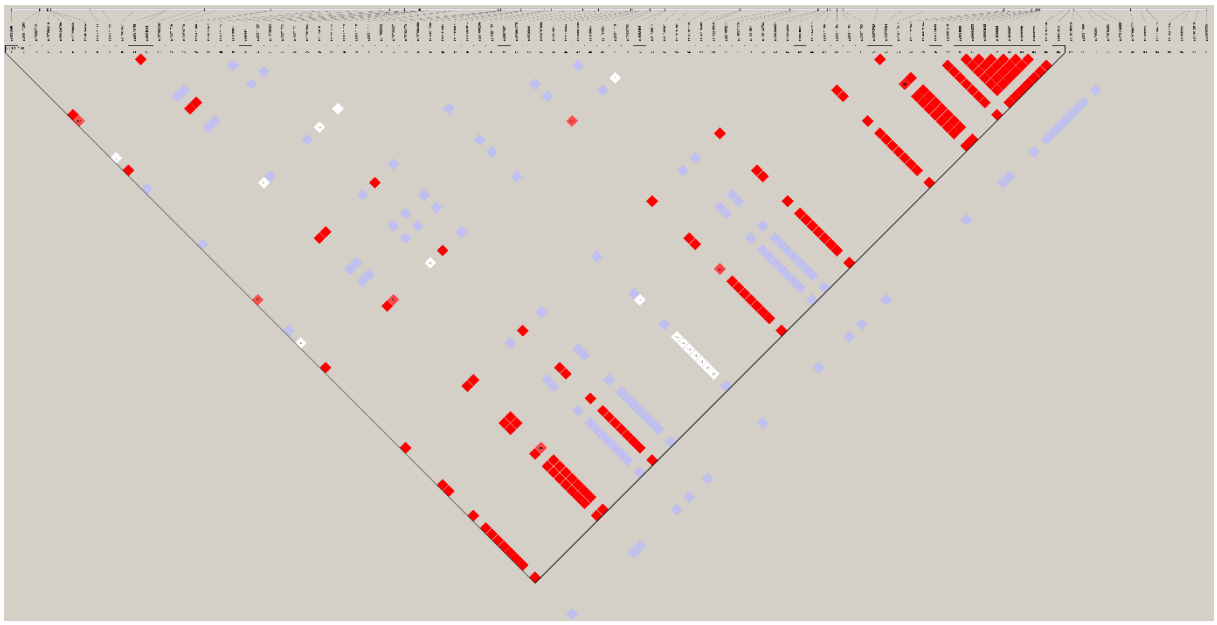
IBS



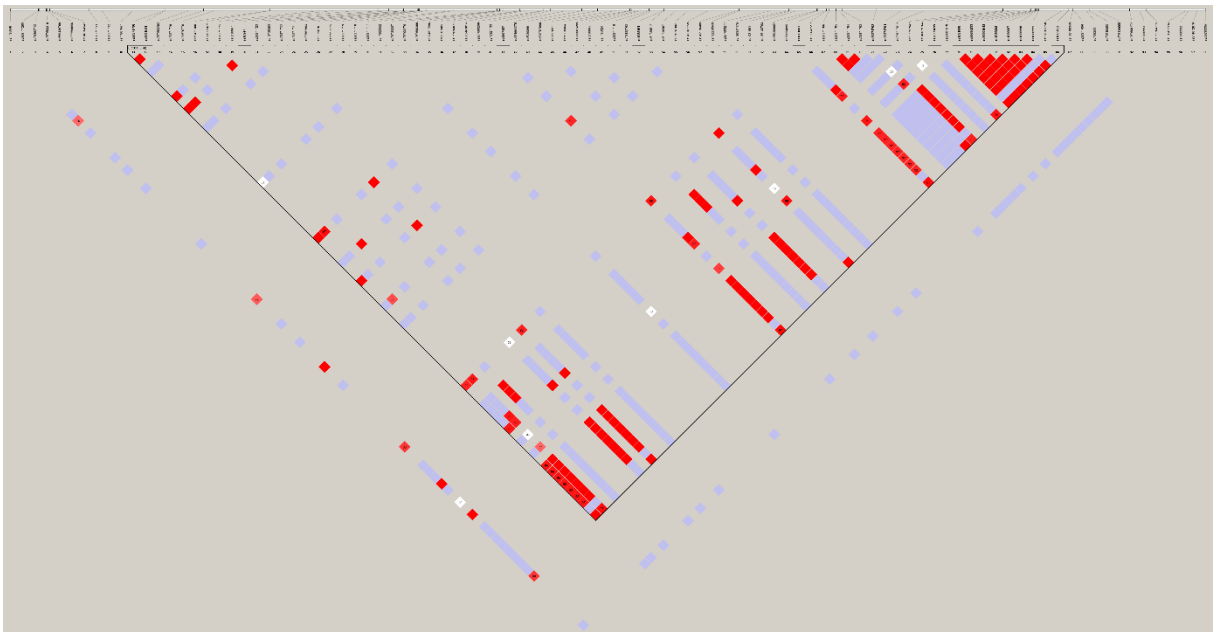
ITU



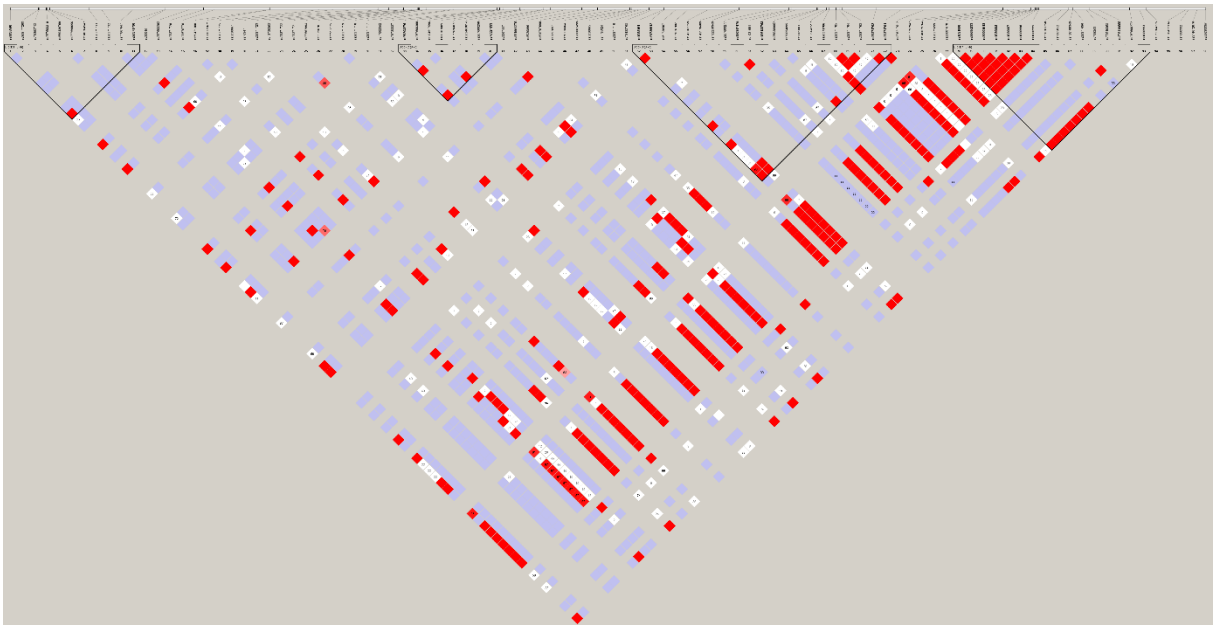
JPT



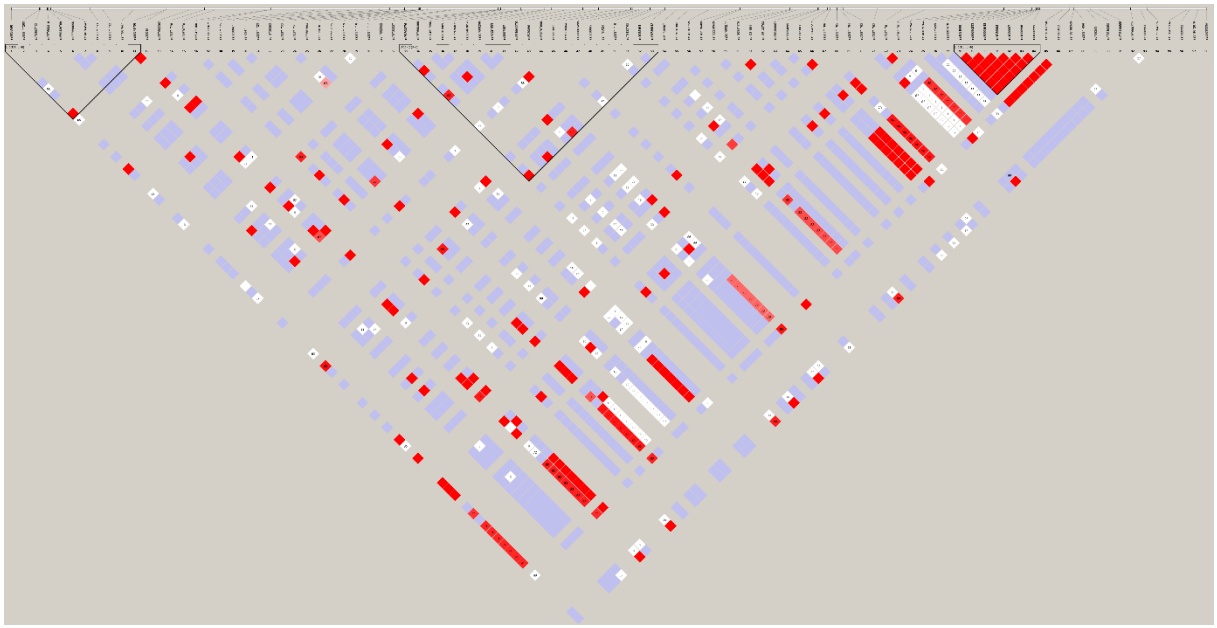
KHV



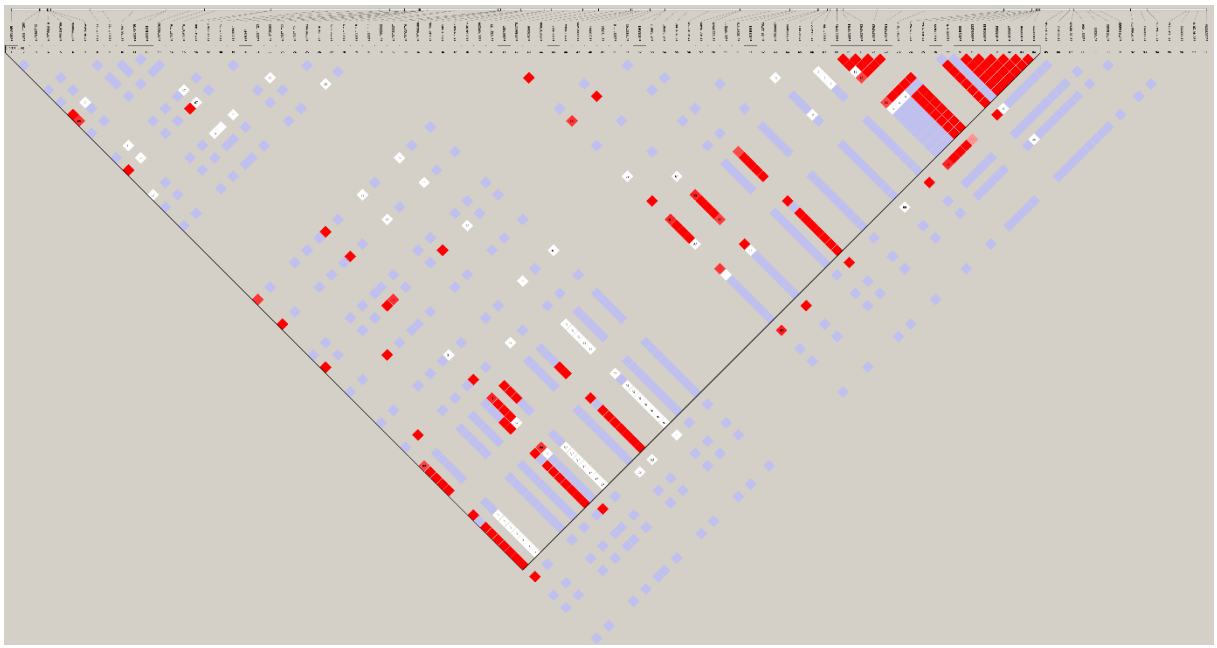
LWK



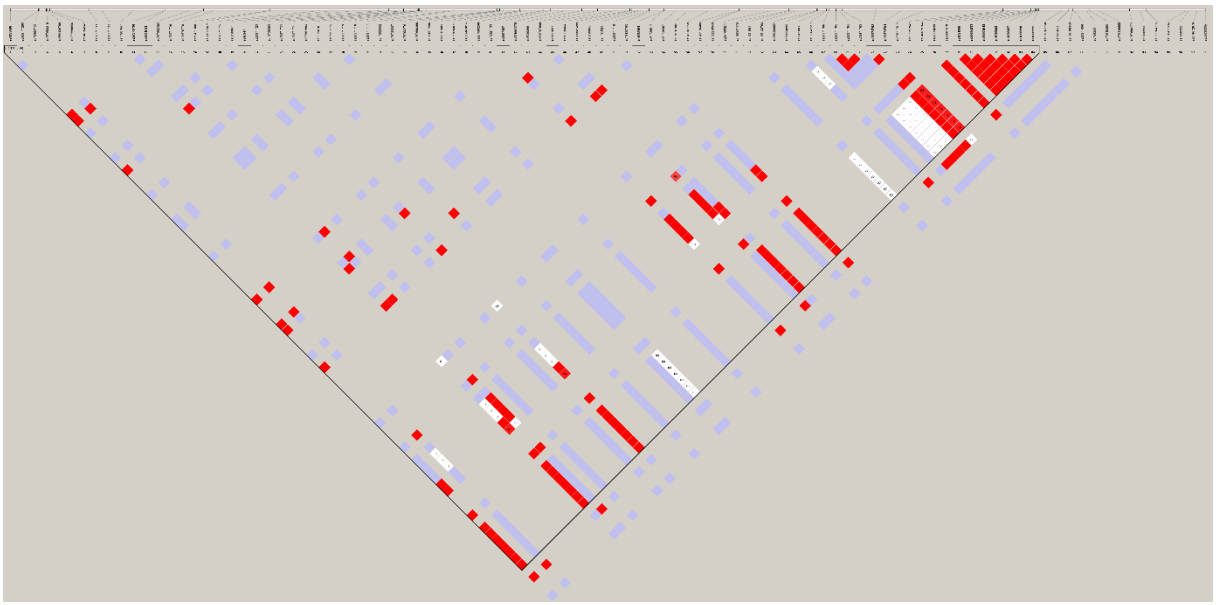
MSL



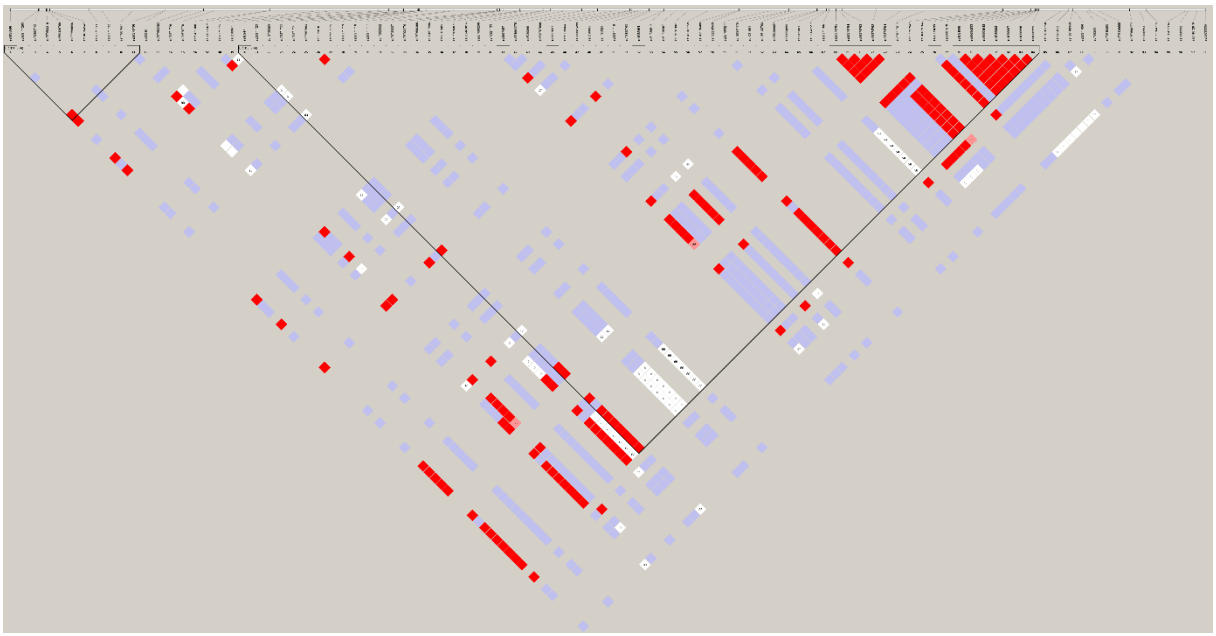
MXL



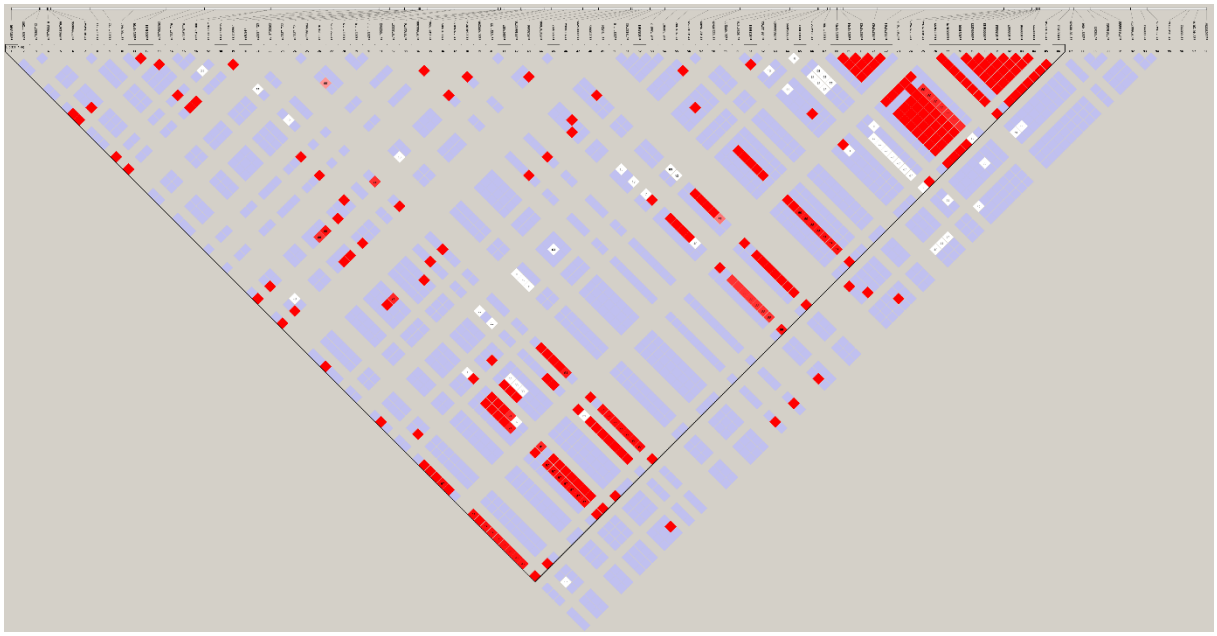
PEL



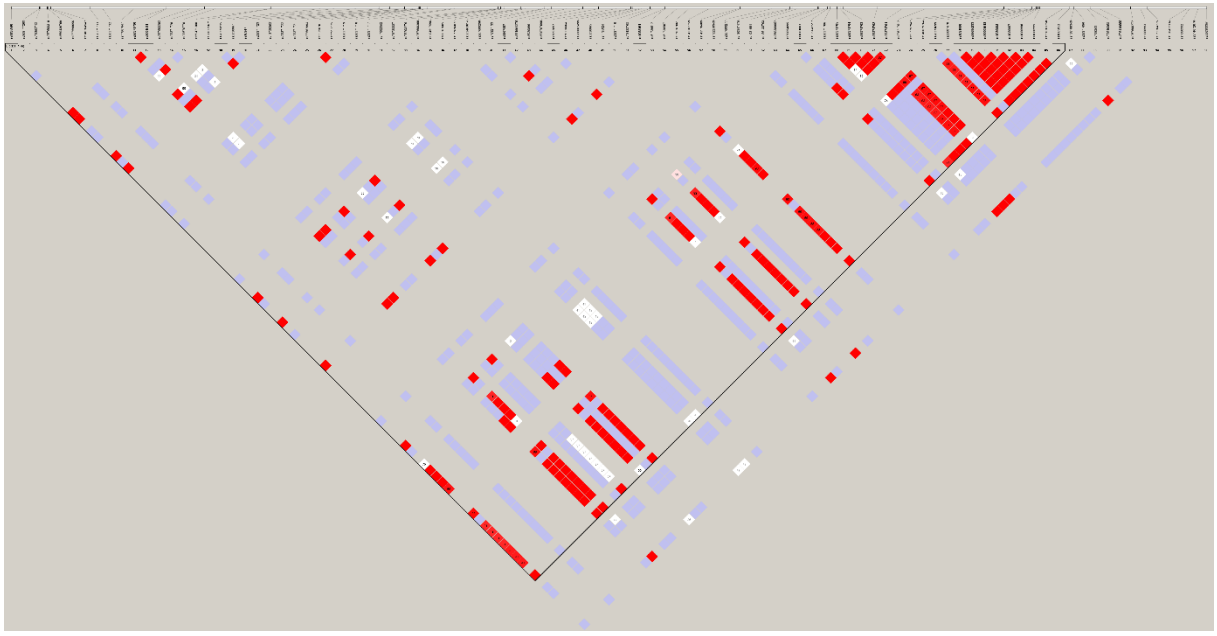
PJL



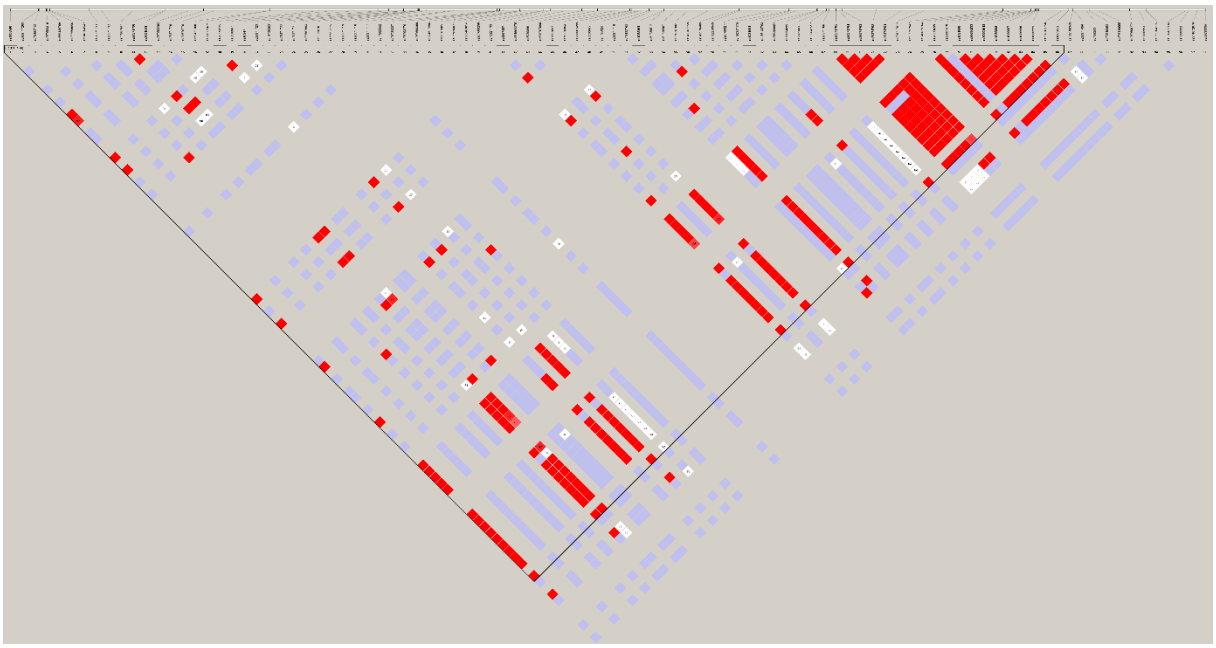
PUR



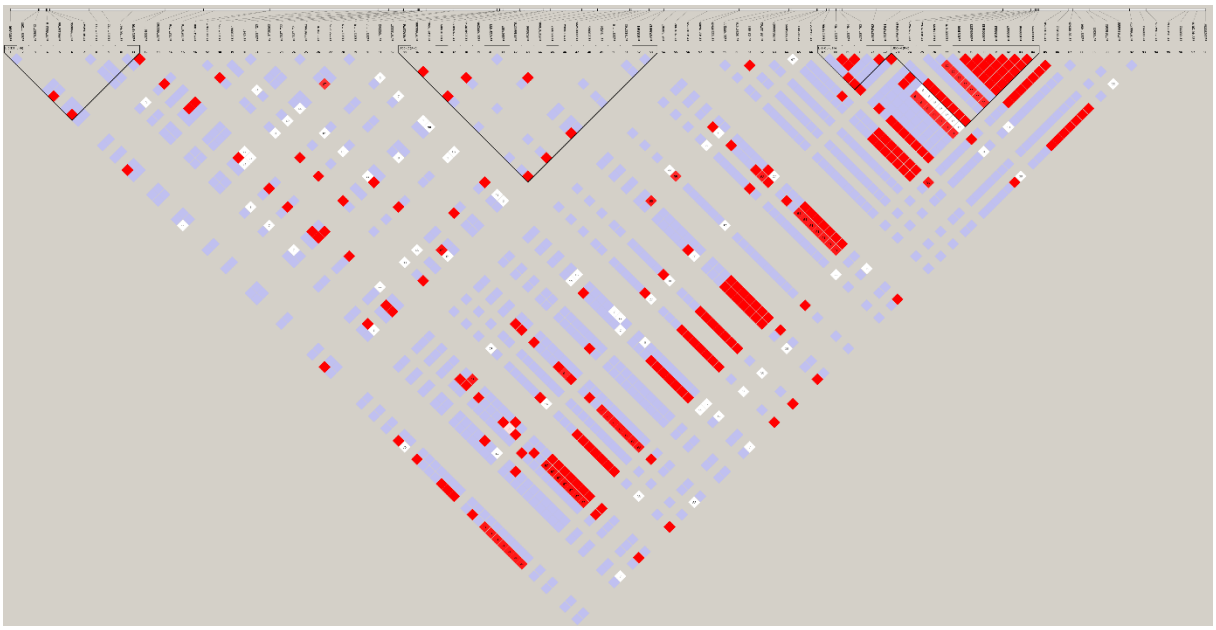
STU



TSI



YRI



3. DISCUSSION

The growing number of population data on genes responsible for absorption, distribution, metabolism and excretion (ADME genes) has stimulated research in the field of population pharmacogenomics. The diversity of genetic polymorphisms can be affected by ancestry, resulting in variation in drug responses among populations (Nagaraj & Toombs, 2021). One of the most studied ADME genes is *CYP2D6* which was the focus of this thesis.

To assess the variability of the *CYP2D6* gene on a global scale, 26 world populations from the publicly available 1000 Genomes database were analyzed (1000 Genomes Project et al., 2015). Within the sequence of the entire gene (ENSG00000100197 chromosome 22: 42,126,499–42,130,881, GRCh38.p12), 86 SNPs were found with a frequency greater than 0.1%. 232 unique haplotypes were reconstructed from polymorphic SNPs, of which 58 were used for haplotype-based PCA analysis, which showed cluster separation for African and East Asian populations, while South Asian, European and American populations were less separated. The separation of African populations is evident in various genetic studies and follows their genetic history (Jing Li et al., 2011; Prohaska et al., 2019). Clustering of East Asian populations is also evident from studies of other pharmacogenes. Li and colleagues showed that ADME genes have very different patterns of population differentiation at global and regional scales (Jing Li et al., 2011). The genomic diversity of modern populations reflects past demographic and evolutionary events, and genetic specificity is especially pronounced in isolated populations where gene exchange with other populations is minimal (Dlouhá et al., 2020; Li, Lao, et al., 2014; Moya et al., 2016; Stojanović Marković, Zajc Petranović, Škobalj, et al., 2022; Weber et al., 2015; Wen et al., 2022; Zajc Petranovic et al., 2019).

One of such isolated populations is the Roma population, an example of a founder population that remained in centuries-long isolation. Due to the complex history of migration and endogamy limited to certain Roma groups, the Roma today are a highly structured population (Chaix et al., 2004; Fraser, 1992). Mitochondrial DNA (mtDNA) research has revealed a clear division between Vlax Roma and other Roma populations that migrated to Europe during the first migration wave (Mendizabal et al., 2011; Salihovic et al., 2011). Similar results were obtained by analyzing autosomal (Gusmao et al., 2010) and Y STR loci (Klaric et al., 2009). The demographic history of the Roma population has put them in the focus of pharmacogenomic research in the last few years (N. Font-Porterias et al., 2021). Previous studies investigated certain SNPs of different genes: *ABCB1* (Sipeky et al., 2011; Zajc Petranovic et al., 2019), *CYP2B6* (Dlouhá et al., 2020; Tomas et al., 2017; Weber et al.,

2015), *NAT* (Stojanović Marković, Zajc Petranović, Škobalj, et al., 2022; Teixeira et al., 2015), *CYP2C19* (Petrović et al., 2019; Sipeky et al., 2013; Teixeira et al., 2015; Zajc Petranovic et al., 2018). Regarding *CYP2D6,* we investigated the entire gene, while prior research focused on a few selected SNPs in the Roma population, such as rs3892097 and rs1065852 (Dlouhá et al., 2020; Weber et al., 2015), or rs35742686 and rs3892097 (Petrović et al., 2019).

The aim of this doctoral dissertation was to determine how the demographic history of the Roma affects the distribution of SNP variants within the *CYP2D6* gene. In order to get an answer to that question, the entire *CYP2D6* gene was sequenced in three Croatian Roma populations using the Genotyping-in-Thousands by sequencing (GT-seq) method (Campbell et al., 2015), which found 43 polymorphic loci. The number of polymorphic loci is similar to that in other investigated populations (Stojanovic Markovic et al., 2022), but their distribution is different. Some of the polymorphic SNPs in the Roma population are monomorphic in publicly available databases. Examples of such loci are: rs368389952, rs566383351, rs374672076, and rs17002852. However, population-specific research such as that of Ahmed and colleagues found two of these loci, rs566383351 (14.59%) and rs374672076 (12.7%), polymorphic in the Pakistani population as well (Ahmed et al., 2018). These results indicate the importance of researching specific populations that, due to their genetic history, do not follow the patterns of large population groups. Similarity of Croatian Roma with European and South Asian populations can be seen in the frequencies of loci rs4987144, rs28371730, rs28371725, rs16947, rs2267447, rs3892097 and rs1065852. This similarity might not be unusual because the migration of Roma from India to Europe started almost 1000 years ago. For rs4987144, rs28371730 there is no determined drug response as for now, and they do not define core alleles. Rs28371725 is an intron variant that defines multiple core alleles (**32, *41, *69, *91,* etc.) and is associated with deutetrabenazine, tamoxifen, and tramadol response. Rs16947 is a missense variant that encodes a myriad of core alleles (**2, *8, *11, *12,* etc.) and is associated with Tamoxifen response, deitetrabenazine response, and ultrarapid metabolism of debrisoquine. Rs2267447 is an intron variant that encodes multiple core alleles (**10, *36, *37,* etc.) and is associated with tramadol response. Rs3892097 is a splice acceptor variant that encodes core allele **4* and is associated with response to many drugs such as debrisoquine, deutetrabenazone, tamoxifen, tramadol, etc. Rs1065852 is a missense variant that encodes multiple core alleles (**10, *36, *37,* etc.) and is associated with response to debrisoquine, deutetrabenazone, tamoxifen, and tramadol.

Given that the Roma gene pool is an admixture of Eurasian populations, the paper of Stojanović Marković and colleagues attempted to determine which SNPs within haplotypes most influence differences among world populations and whether these loci have a translational role (Stojanović Marković, Zajc Petranović, Škarić-Jurić, et al., 2022).

Using the locus-by-locus AMOVA method, 11 SNPs were found whose allele frequency differences contribute the most to differences between populations.

SNPs characteristic of populations of the African continental group are almost completely absent in other world populations, including in Croatian Roma populations. The five SNPs (rs75203276, rs59421388, rs61736512, rs76327133, and rs80262685) have a similar frequency within all African populations (8-19%), except for the population of African Ancestry in the Southwest USA, where the frequency is 3-5%. Rs28371706, whose alternative allele A causes a missense variant, is responsible for defining the star alleles **17, *40, *58, *64, *82, *141* and **154*. According to Lymperopoulos and colleagues (Lymperopoulos et al., 2015) and Masimirembwa and colleagues (Masimirembwa et al., 1996), *CYP2D6*17* occurs in at least 30% of Africans, is associated with reduced enzyme activity, and individuals carrying this allele are classified as intermediate metabolizers (IM). According to the ClinVar, rs28371706 is associated with response to tamoxifen and deutetrabenazine.

East Asian populations are characterized by four SNPs (rs2267447, rs1065852, rs2004511, and rs1081003) and their frequency in Vietnamese and three Chinese populations is 60-68%, which means that their minor allele has a higher frequency than the major allele, which makes the less frequent major allele a minor allele in these populations. The distribution of these SNPs is somewhat different in the Japanese population and ranges from 36-39%. The deviation of the values of the Japanese population is not surprising given their relative isolation throughout history. These SNPs also appear in other world populations, mostly in the 11-26% range, except for rs1081003, which is far less represented in other world populations and exceeds a frequency of 10% only in the populations of Yoruba, Sierra Leone, Italy, Telugu Indians and in the population of Bengalis from Bangladesh.

When we compare the values of these four SNPs from the East Asian group with the values from the research of Croatian Roma populations, we see that the frequencies of rs1081003 (1-6%) in Croatian Roma are similar to those in European and South Asian populations (1-6%), rs2267447 and rs1065852 have ranges (23-28%) similar to some

European populations (Great Britain, Italy, and Utah population originating from Central Europe) and Bengalis from Bangladesh, while rs2004511 shows interesting results with frequencies similar to those for Japan (34-40%), which is perhaps not so surprising since we know that Roma are also an isolated population, and it is probably their isolation is the cause of the increased frequency. The association of the alternative allele C of rs2004511 was recorded in the ClinVar with the response to tramadol. In the PharmVar, we found that this allele defines suballeles for a number of star alleles (**4*, **10*, **36*, **39*, etc.). López-García and colleagues (López-García et al., 2017) found that SNP rs1065852 may influence the efficacy of antiepileptic drugs because it is involved in the star allele **4*. Runcharoen and colleagues (Runcharoen et al., 2021) found that this SNP is in high LD with rs1081003 and rs1065852 in the populations of the Philippines, Thailand, Vietnam and Laos, and that mutations in these key SNPs defining star alleles **10* and **54* cause reduced CYP2D6 enzyme activity.

The distribution of rs16947 is the most interesting and occurs in all studied world populations, as well as in three populations of Croatian Roma. This SNP is most characteristic of African populations with a MAF frequency of 42-65% (50-65% if we exclude US populations of South West African origin). The MAF range in East Asian populations is 13-17%, South Asian 25-45%, and in European populations 31-38%. The MAF range in the population of Croatian Roma is 33-43%, which places it in the range of the populations of South Asia and Europe, which is the expected result due to the migrations of Roma that started from South Asia to Europe.

Using Phase software (Stephens & Donnelly, 2003; Stephens et al., 2001) we were able to reconstruct 93 haplotypes from polymorphic loci. Intra-population analysis based on haplotypes showed the highest diversity rate in the Roma population from Baranja, and the lowest diversity rate in Medjimurje. The least number of haplotypes (37) were found in the population of Roma from Medjimurje, which is in accordance with previous genetic analyses which established that they are the least diverse Roma population in Croatia (Klaric et al., 2009; Salihovic et al., 2011). A direct comparison of the haplotypes of the Roma population and haplotypes from the researched world populations is not possible. Namely, *CYP2D6* gene is surrounded with two pseudogenes, has repetitive sequences, copy number variation and high density of polymorphisms. All this makes it prone to errors when used short-read sequencing if amplicons are not specific enough. This is pronounced when this gene is analysed as part of a whole genome sequence and can result in variant call differences. Such

example in here analysed data are SNPs rs1985842 and rs4987144 which are monomorphic in 1000 genomes database (WGS using short-red method) but polymorphic in Croatian Roma samples and gnomAD database.

In order to evaluate the pharmacogenetic aspect of the Roma population, 89 reconstructed haplotypes were translated into CYP alleles according to the available CYP nomenclature (www.pharmvar.org). The alleles with the highest frequency in the three Roma populations were compared with the frequencies of the same star alleles weighted by population size in populations worldwide, grouped by ethnicity. The results showed that the most represented haplotype in all three Roma populations is *CYP2D6*1*. This haplotype is also considered the reference and constitutes the majority of all haplotypes in Europe and Africa. Of the three Roma populations, the lowest frequency was observed among Balkan Roma (25.91%). The second most common haplotype in Balkan Roma (25.39%) and Roma from Medjimurje (27.10%) is *CYP2D6*2*. Comparing with the data of (Gaedigk et al., 2017), the frequency of *CYP2D6*2* in these two Roma populations is similar to that of European and South Asian populations. These results are consistent with the findings of (Naveen et al., 2006), which showed a similar distribution of this haplotype among the populations of Europe and South Asia. Sistonen and colleagues hypothesized that long-term selective pressure maintains a high frequency of haplotypes coding for a fully functional enzyme, resulting in a uniform geographic distribution of **1* and **2* alleles (Sistonen et al., 2007).

The most common non-functional allele is *CYP2D6*4*. A single base polymorphism at the splice site of intron 3 and exon 4 leads to the development of a defective variant of the enzyme (rs3892097; c.1846G>A), which is also called the *CYP2D6*4* isoform and results in a slow metabolizer (Gaedigk, Dinh, et al., 2018; Sistonen et al., 2007). This allele is the most frequent in European populations (18%). Among Roma populations, the highest frequencies were observed in Balkan Roma (24.87%) and Baranja Roma (20.51%), which were higher than in European populations. These results are consistent with the results of research on Hungarian Roma, Czech Roma and South Asian populations (Dlouhá et al., 2020; Weber et al., 2015). The *CYP2D6*10* reduced function allele is an intermediate metabolizer. Individuals who are homozygous for this allele are at risk of adverse drug reactions, but less so than those who are poor metabolizers (Zanger et al., 2004). This allele is most frequent in East and South Asian populations (9-44%), and least frequent in European populations (<2%) (Gaedigk et al., 2017; Pratt et al., 2021). The Roma from Medjimurje show a frequency of **10* alelle (10%) close to the populations of South Asia, especially Indians from South India

(Naveen et al., 2006). Another allele of reduced function, *CYP2D6*41*, was found in Croatian Roma. Roma from Medjimurje, with a frequency of 6%, are most similar to South Asian populations (9.9%) and European populations (8.5%), while Balkan (17%) and Baranja Roma (18%) are closer to Middle Eastern populations (17%). The **41* frequency in Baranja Roma is also the highest frequency in world.

Comparing the distribution of the five most common star alleles in the three Roma populations, no significant difference was observed between the Baranja and Balkan Roma. However, the Roma from Medjimurje significantly differ from the other two Roma populations, which is in accordance with the research of Salihović and colleagues, where the highest rate of isolation was shown among the Roma from Medjimurje compared to the other two populations (Salihovic et al., 2011). Overall, looking at the results of the translation of haplotypes into star alleles, it can be seen that the distribution of *CYP2D6* slow alleles (**9, *10, *17, *29, *45-46*) and null function alleles (**4, *5, *6*) differs by continents, probably due to demographic events (Sistonen et al., 2007). Individuals who are slow metabolizers may be better (or worse) metabolizers of certain classes of chemical compounds. This is not necessarily a bad thing as it may reduce the risk of side effects if the toxic compound is activated by CYP2D6 enzyme metabolism (Fuselli et al., 2010). The CYP2D6 substrates in humans include a large number of common medication, drugs, exogenous substances (e.g. alkaloids, herbicides) and some endogenous compounds (e.g. progesterone and estrogen) (Wang et al., 2009).

Natural selection has affected many ADME genes and left specific signals in patterns of genetic diversity around selected loci (Nielsen et al., 2007). A possible explanation is that different environments and diets after migration “out of Africa” could exert a strong selection pressure on ADME genes since they are mainly involved in defense against xenobiotics (Jing Li et al., 2011).

Patterns of *CYP2D6* haplotype blocks that vary among continental groups have been used to study population variation. The most homogeneous pattern is in Europe, followed by America, South and East Asia, while the most diverse pattern is in African populations. The studied blocks of haplotypes encompass the region of SNPs with strong LD as a consequence of lack of recombination, and are based on confidence intervals (Gabriel et al., 2002). Consistent with their genetic history, the largest number of blocks are present in African populations due to relatively large effective population sizes over long periods of time. We observed a haplotype block of the *CYP2D6* gene in the range rs1081000-rs1080995 (r^2>0.8)

in all African populations (except Yoruba). As it was explained earlier, direct comparison of Croatian Roma and populations from 1000 genomes was not possible so that also applied to the comparison of LD within *CYP2D6* gene of Croatian Roma and other populations they had contact with on their migration routes. Within Croatian Roma, distribution of LD was similar between Baranja and Balkan Roma , but was slightly different in Roma from Medjimurje. It would be interesting to find out which haplotypes in Roma population are ancestral to proto-Roma Indian population and which are the result of admixture for surrounding populations. However, for precise estimations it would be necessary to have highly comparable haplotypes, preferably generated using third-generation long-read *CYP2D6* sequencing (Fukunaga et al., 2020) because that is a method that can overcome shortcomings of short-read sequencing.

The complete metabolizing phenotype is obtained by translation of diplotypes (combination of a person`s two *CYP2D6* alleles). In three Croatian Roma populations, 28 star diplotypes were found, which were translated into phenotypes and sorted into three groups of metabolizers: normal, medium and slow. The most common was a normal metabolizer phenotype (51.6%-65.1%), followed by an intermediate metabolizer phenotype (34.9%-41.1%), and a slow metabolizer was the rarest (1.7-7.4%), and was not found in the Roma population from Medjimurje. Slow metabolizers were also not observed in the Hungarian Roma population (Weber et al., 2015). However, we should take this grouping of metabolizers with caution because it is based only on the frequency of polymorphisms without taking into account possible changes in gene copy number that are frequent for this locus and significantly affect the phenotype (Pratt et al., 2021).

The phenotype depends not only on the gene sequence but also on the regulatory regions. Understanding the variety of regulatory components that influence gene expression can help explain unexplained variability in gene activity. How enhancer/promoter activity and variants in the *CYP2D6* gene impact metabolism of drugs has been poorly investigated. So far, only a small number of regulatory variants of the *CYP2D6* gene have been investigated and this research has mostly focused on two fully linked loci; rs133333 (G > A) and rs5758550 (G > A), classified as enhancers and located 116 kb downstream of the gene (Wang et al., 2014).

In order to evaluate the variation in the regulatory region of the *CYP2D6* gene among: (a) the three Croatian Roma populations, and (b) between Croatian Roma populations and populations of Europe and Asia from the publicly available database of 1000 genomes; 16

loci from the promoter region of the *CYP2D6* gene and 2 enhancers (rs133333 and rs5758550) were genotyped.

Although differences in the distribution of alleles and genotypes in the coding region were observed in the Croatian Roma population (Stojanovic Markovic et al., 2022), this was not the case with the regulatory region. Linkage disequilibrium (LD) analyses were performed to elucidate the relationship between SNPs in the regulatory region and star allele defining SNPs from the coding region. Significant LDs between SNPs in the regulatory and gene regions may affect *CYP2D6* transcription and consequently drug metabolism. So far, the interaction of the rs5758550 SNP has been studied the most (Elias et al., 2020). This SNP was in high LD (r^2>0.8) only with SNPs rs133333 and rs1080985 in all three Roma populations. High LD in world populations was observed in interaction with SNP rs16947 in Finnish population. Rs1080985 was linked with increased expression of CYP2D6 enzyme in the human liver (Raimundo et al., 2000; Zanger et al., 2001), but today we believe that this SNP has no effect on the body's function (https://www.ncbi.nlm. nih.gov/clinvar/ (accessed 15.03.2022)). Wang and colleagues suggested that the elevated mRNA expression associated with this SNP could be explained by LD between the rs5758550/rs133333 enhancer SNP and rs1080985 (Wang et al., 2014). The reconstructed haplotypes in Croatian Roma have allele G at rs5758550 and allele C at rs1080985 on more than 20% of the chromosomes.

LDs between SNPs of the regulatory region and SNPs defining main star alleles were examined. SNPs rs1080989/rs28588594 have high LD with allele **4* (rs3892097), but r^2 value greater than 0.8 was found only in Balkan Roma. The non-functional *CYP2D6*4* allele most often found in European populations (18%), had a higher frequency in Balkan and Baranja Roma, while in Medjimurje Roma it was lower than in European populations, but still above the frequencies of other world populations. LD r^2 value greater than 0.8 was found in all Roma populations between rs1080989/rs28588594 and allele **10* (rs1065852). *CYP2D6*10* is an allele of reduced functionality most often found in East and South Asian populations (9-44%). In the Roma population, the frequency was 6%, and in European populations, it was less than 2%. Combinations of *CYP2D6* SNPs from the regulatory and coding regions that were in high LD in Roma populations were also present in most world populations, but we noticed differences within Roma populations. An example of this are differences in LD values between groups at *CYP2D6*4* in relation to rs1080989/rs28588594.

Multidimensional scaling plots (MDS) (n=7) were constructed, and this method can be used to discover hidden patterns in the correlation matrix. Most MDS graphs showed a clear

separation of East Asian populations from other world populations. This separation was also recorded in other studies related to ADME genes (Jing Li et al., 2011; Skaric-Juric et al., 2018). The Balkan Roma population was remotely positioned in comparison to other populations on the MDS plots, while the Baranja Roma population was positioned in the vicinity of South Asian populations, and the Medjimurje Roma population was positioned either with South Asian populations (rs5758550, rs28624811 and rs1080985) or close to European populations (rs28735595, rs28588594 and rs1080989).

Numerous studies focused on population variation of the ADME gene, but isolated populations are still not sufficiently represented in these studies (Zhang et al., 2019). Since genetic variation within this group of genes mainly occurs due to human adaptation to the environment (Fuselli, 2019), there are large genetic differences in ADME genes between modern populations (Li, Lou, et al., 2014; Maisano Delser & Fuselli, 2013). These differences are particularly accentuated in minority (Li, Lou, et al., 2014) and isolated populations (Zhou & Lauschke, 2018) with specific demographic histories.

Due to their socio-economic status, Roma have difficult access to health care and an increased risk of diseases such as diabetes, heart disease, etc. (Zajc Petranović et al., 2021). Given that the CYP2D6 is involved in the metabolism of many widely available drugs (Gardiner & Begg, 2006), it is of great interest for research in minority and isolated populations. For these reasons, it is important to include isolated, and often neglected, populations such as the Roma in research in order to gain knowledge about their specificities that should be taken into account during treatment. To our knowledge, this study is the most comprehensive study of the *CYP2D6* gene in Roma populations, and thus aims to contribute to this goal.

This study confirmed the South Asian origin of Croatian Roma, but also showed the importance of researching Roma populations as separate groups due to the significant differences that exist between them as a result of their different history, culture and environment. These differences are most likely caused by genetic drift resulting from endogamy within the studied Roma groups.

4. CONCLUSIONS

Analyses of *CYP2D6* gene and promotor regions in Croatian Roma and world populations revealed the following conclusions:

- Among 26 world-wide populations 89 polymorphic SNPs with frequency higher than 0.1% was found in *CYP2D6* gene
- 232 haplotypes were reconstructed from polymorphic SNPs in gene region
- The eight most common haplotypes account for 74% of all haplotypes worldwide all populations share the most common haplotype determining the star allele **1*.
- The highest interpopulation differentiation is among East Asian populations and the lowest among European populations
- Locus-by-locus analyses revealed 11 SNP loci substantially affecting inter-population differentiation, six of which are specific to African (rs75203276, rs59421388, rs61736512, rs76327133, rs80262685, and rs28371706), four to East Asian populations (rs2267447, rs1065852, rs2004511 and rs1081003), while one is present globally (rs16947). Five of these SNPs (rs2004511, rs1065852, rs2267447, rs28371706, and rs16947) contribute to the known pharmacogenomic effects on clinical outcomes of drugs metabolized by CYP2D6.
- In Croatian Roma population 43 polymorphic SNPs were found. SNPs found to be polymorphic in the investigated Roma groups had higher MAFs in South Asian than in the European populations.
- 93 *CYP2D6* haplotypes were reconstructed from polymorphic SNPs, the highest diversity was found in the population of Roma from Baranja and the lowest among Roma from Medjimurje.
- Allele *CYP2D6*1* was the most common in each of the three Roma groups individually, and in the entire Roma sample. The other most prevalent alleles were **2*, **4*, **10*, and **41*.
- The normal metabolizing phenotype was found in 57.9% of examinees, the intermediate metabolizing phenotype in 39.3%, and the poor metabolizing phenotype in 2.8% of the Croatian Roma.
- 10 out of 18 loci were polymorphic in the regulatory region of *CYP2D6* in Croatian Roma Population
- High LD was found between SNP rs16947 which defines star allele **2* and promoter SNPs rs1080983 and rs28624811, between SNP rs3892097 which defines star allele **4* and regulatory region SNPs rs1080989 and rs28588594 and between SNP

rs1065852 which defines decreased function allele **10* and regulatory region SNPs rs1080989 and rs28588594.

- An overall comparison of the analyzed LD values revealed the greatest variety in the populations of East Asia and uniform but distinct distribution in populations of Europe and South Asia

In summary, demographic history, migrations and endogamy of Croatian Roma populations influenced the distribution of variants within *CYP2D6* gene. Genetic drift that operates in Croatian Roma populations led to the accumulation of globally rare variants. South Asian origin of Croatian Roma can be seen in the frequencies of polymorphic variants on many SNPs, as well as in elevated frequencies of star alleles **10* and **41*. The *CYP2D6* gene's impaired-function star alleles **2* and **4*, and higher LD values between the investigated SNPs in the promoter region and their star-defining alleles could be used in Roma to increase genotyping effectiveness. The three socio-culturally different Roma groups studied differ significantly in the distribution of star alleles, which confirms the importance of studying different Roma groups separately.

5. REFERENCES

1000 Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. https://doi.org/10.1038/nature15393

Abrahams, E. (2009). Letter from the Executive Director: Latest News & Updates from the Personalized Medicine Coalition. *Per Med*, *6*(3), 245. https://doi.org/10.2217/pme.09.16

Ahituv, N. (2012). Gene Regulatory Elements. In *Gene Regulatory Sequences and Human Disease* (pp. 1-17). https://doi.org/10.1007/978-1-4614-1683-8_1

Ahmed, S., Zhou, J., Zhou, Z., & Chen, S.-Q. (2018). Genetic Polymorphisms and In Silico Mutagenesis Analyses of CYP2C9, CYP2D6, and CYPOR Genes in the Pakistani Population. *Genes*, *9*(10). https://doi.org/10.3390/genes9100514

Aklillu, E., Carrillo, J. A., Makonnen, E., Hellman, K., Pitarque, M., Bertilsson, L., & Ingelman-Sundberg, M. (2003). Genetic Polymorphism of CYP1A2 in Ethiopians Affecting Induction and Expression: Characterization of Novel Haplotypes with Single-Nucleotide Polymorphisms in Intron 1. *Molecular Pharmacology*, *64*(3), 659-669. https://doi.org/10.1124/mol.64.3.659

Alzahrani, A. M., & Rajendran, P. (2020). The Multifarious Link between Cytochrome P450s and Cancer. *Oxidative Medicine and Cellular Longevity*, *2020*, 1-18. https://doi.org/10.1155/2020/3028387

Augustini ab Hortis, S. (1775). *Von dem heutigen Zustande, sonderbaren Sitten und Lebensart, wie auch von denen übrigen Eigenschaften und Umständen der Zigeuner in Ungarn* [On the present situation, special manners and way of life, as well as other characteristics and gifts of the Gypsies in Hungary]. Kaiserlich Königliche allergnädigste privilegierte Anzeigen aus sämstlichen Kaiserl. Königl. Erbläder.

Beoris, M., Amos Wilson, J., Garces, J. A., & Lukowiak, A. A. (2016). CYP2D6 copy number distribution in the US population. *Pharmacogenet Genomics*, *26*(2), 96-99. https://doi.org/10.1097/fpc.0000000000000188

Berm, E. J., Looff, M.d, Wilffert, B., Boersma, C., Annemans, L., Vegter, S., Boven, J. F., & Postma, M. J. (2016). Economic Evaluations of Pharmacogenetic and Pharmacogenomic Screening Tests: A Systematic Review. Second Update of the Literature. *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0146262

Bernauer, U., Heinrich-Hirsch, B., Tönnies, M., Peter-Matthias, W., & Gundert-Remy, U. (2006). Characterisation of the xenobiotic-metabolizing Cytochrome P450 expression pattern in human lung tissue by immunochemical and activity determination. *Toxicology Letters*, *164*(3), 278-288. https://doi.org/10.1016/j.toxlet.2006.01.007

Bertz, R. J., & Granneman, G. R. (1997). Use of In Vitro and In Vivo Data to Estimate the Likelihood of Metabolic Pharmacokinetic Interactions. *Clinical Pharmacokinetics*, *32*(3), 210-258. https://doi.org/10.2165/00003088-199732030-00004

Bolwell, G. P., Bozak, K., & Zimmerlin, A. (1994). Plant cytochrome p450. *Phytochemistry*, *37*(6), 1491-1506. https://doi.org/10.1016/s0031-9422(00)89567-9

Božina, N. (2017). Farmakogenomika i farmakovigilancija. *Medicus*, *26*(1 Farmakovigilancija), 13-22. https://hrcak.srce.hr/185266

Butler, J. E. F., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development*, *16*(20), 2583-2592. https://doi.org/10.1101/gad.1026202

Calhoun, V. C., Stathopoulos, A., & Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences*, *99*(14), 9243-9247. https://doi.org/10.1073/pnas.142291299

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855-867. https://doi.org/10.1111/1755-0998.12357

Cascorbi, I., Bruhn, O., & Werk, A. N. (2013). Challenges in pharmacogenetics. *Eur J Clin Pharmacol*, *69*(S1), 17-23. https://doi.org/10.1007/s00228-013-1492-x

Chaix, R., Austerlitz, F., Morar, B., Kalaydjieva, L., & Heyer, E. (2004). Vlax Roma history: what do coalescent-based methods tell us? *Eur J Hum Genet*, *12*(4), 285-292. https://doi.org/10.1038/sj.ejhg.5201126

Chen, W., Lee, M.-K., Jefcoate, C., Kim, S.-C., Chen, F., & Yu, J.-H. (2014). Fungal Cytochrome P450 Monooxygenases: Their Distribution, Structure, Functions, Family Expansion, and Evolutionary Origin. *Genome Biology and Evolution*, *6*(7), 1620-1634. https://doi.org/10.1093/gbe/evu132

Chinta, S. J., Pai, H. V., Upadhya, S. C., Boyd, M. R., & Ravindranath, V. (2002). Constitutive expression and localization of the major drug metabolizing enzyme, cytochrome P4502D in human brain. *Molecular Brain Research*, *103*(1-2), 49-61. https://doi.org/10.1016/s0169-328x(02)00177-8

Cooper, D. Y., Levin, S., Narasimhulu, S., Rosenthal, O., & Estabrook, R. W. (1965). Photochemical Action Spectrum of the Terminal Oxidase of Mixed Function Oxidase Systems. *Science*, *147*(3656), 400-402. https://doi.org/10.1126/science.147.3656.400

Dalen, P., Frengell, C., Dahl, M. L., & Sjoqvist, F. (1997). Quick onset of severe abdominal pain after codeine in an ultrarapid metabolizer of debrisoquine. *Ther Drug Monit*, *19*(5), 543-544. https://doi.org/10.1097/00007691-199710000-00011

Danielson, P. B., Bentham Science Publisher. (2002). The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans. *Current Drug Metabolism*, *3*(6), 561-597. https://doi.org/10.2174/1389200023337054

Distlerath, L. M., & Guengerich, F. P. (1984). Characterization of a human liver cytochrome P-450 involved in the oxidation of debrisoquine and other drugs by using antibodies raised to the analogous rat enzyme. *Proceedings of the National Academy of Sciences*, *81*(23), 7348-7352. https://doi.org/10.1073/pnas.81.23.7348

Dlouhá, L., Adámková, V., Šedová, L., Olišarová, V., Hubáček, J. A., & Tóthová, V. (2020). Five genetic polymorphisms of cytochrome P450 enzymes in the Czech non-Roma and Czech Roma population samples. *Drug Metabolism and Personalized Therapy*, *0*(0). https://doi.org/10.1515/dmdi-2020-0103

Durst, F., & O'Keefe, D. P. (1995). Plant Cytochromes P450: An Overview. *Drug Metabolism and Drug Interactions*, *12*(3-4). https://doi.org/10.1515/dmdi.1995.12.3-4.171

Eichelbaum, M., Baur, M. P., Dengler, H. J., Osikowska-Evers, B. O., Tieves, G., Zekorn, C., & Rittner, C. (1987). Chromosomal assignment of human cytochrome P-450 (debrisoquine/sparteine type) to chromosome 22. *British Journal of Clinical Pharmacology*, *23*(4), 455-458. https://doi.org/10.1111/j.1365-2125.1987.tb03075.x

Elias, A. B. R., Araujo, G. S., de Souza, S. J., & Suarez-Kurtz, G. (2020). Distribution and linkage disequilibrium of the enhancer SNP rs5758550 among Latin American populations: influence of continental ancestry. *Pharmacogenet Genomics*, *30*(4), 67-72. https://doi.org/10.1097/FPC.0000000000000398

Estabrook, R. W. (2003). A passion for P450s (rememberances of the early history of research on cytochrome P450). *Drug Metabolism and Disposition*, *31*(12), 1461-1473. https://doi.org/10.1124/dmd.31.12.1461

Estabrook, R. W., Cooper, D. Y., & Rosenthal, O. (1963). The Light Reversible Carbon Monoxide Inhibition of the Steroid C21-Hydroxylase System of the Adrenal Cortex. *Biochem Z*, *338*, 741-755. https://www.ncbi.nlm.nih.gov/pubmed/14087340

Evans, W. E., & Relling, M. V. (1999). Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. *Science*, *286*(5439), 487-491. https://doi.org/10.1126/science.286.5439.487

Fleeman, N., Dundar, Y., Dickson, R., Jorgensen, A., Pushpakom, S., McLeod, C., Pirmohamed, M., & Walley, T. (2010). Cytochrome P450 testing for prescribing antipsychotics in adults with schizophrenia: systematic review and meta-analyses. *The Pharmacogenomics Journal*, *11*(1), 1-14. https://doi.org/10.1038/tpj.2010.73

Font-Porterias, N., Caro-Consuegra, R., Lucas-Sánchez, M., Lopez, M., Giménez, A., Carballo-Mesa, A., Bosch, E., Calafell, F., Quintana-Murci, L., & Comas, D. (2021). The Counteracting Effects of

Demography on Functional Genomic Variation: The Roma Paradigm. *Molecular Biology and Evolution*, *38*(7), 2804-2817. https://doi.org/10.1093/molbev/msab070

Font-Porterias, N., Gimenez, A., Carballo-Mesa, A., Calafell, F., & Comas, D. (2021). Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. *Front Genet*, *12*, 683880. https://doi.org/10.3389/fgene.2021.683880

Fraser, A. (1992). *The Gypsies*. Blackwell Publishers.

Fraser, A., Mondas, A., & Costa, T. (1998). *História do povo cigano*.

Fukunaga, K., Hishinuma, E., Hiratsuka, M., Kato, K., Okusaka, T., Saito, T., Ikeda, M., Yoshida, T., Zembutsu, H., Iwata, N., & Mushiroda, T. (2020). Determination of novel CYP2D6 haplotype using the targeted sequencing followed by the long-read sequencing and the functional characterization in the Japanese population. *J Hum Genet*, *66*(2), 139-149. https://doi.org/10.1038/s10038-020-0815-x

Fuselli, S. (2019). Beyond drugs: the evolution of genes involved in human response to medications. *Proc Biol Sci*, *286*(1913), 20191716. https://doi.org/10.1098/rspb.2019.1716

Fuselli, S., de Filippo, C., Mona, S., Sistonen, J., Fariselli, P., Destro-Bisol, G., Barbujani, G., Bertorelle, G., & Sajantila, A. (2010). Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human CYP2D6 locus. *Pharmacogenet Genomics*, *20*(8), 485-499. https://doi.org/10.1097/FPC.0b013e32833bba25

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*(5576), 2225-2229. https://doi.org/10.1126/science.1069424

Gaedigk, A. (2013). Complexities of CYP2D6 gene analysis and interpretation. *International Review of Psychiatry*, *25*(5), 534-553. https://doi.org/10.3109/09540261.2013.825581

Gaedigk, A., Casey, S. T., Whirl-Carrillo, M., Miller, N. A., & Klein, T. E. (2021). Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. *Clin Pharmacol Ther*, *110*(3), 542-545. https://doi.org/10.1002/cpt.2321

Gaedigk, A., Dinh, J. C., Jeong, H., Prasad, B., & Leeder, J. S. (2018). Ten Years' Experience with the CYP2D6 Activity Score: A Perspective on Future Investigations to Improve Clinical Predictions for Precision Therapeutics. *J Pers Med*, *8*(2). https://doi.org/10.3390/jpm8020015

Gaedigk, A., Ingelman-Sundberg, M., Miller, N. A., Leeder, J. S., Whirl-Carrillo, M., Klein, T. E., & PharmVar Steering, C. (2018). The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin Pharmacol Ther*, *103*(3), 399-401. https://doi.org/10.1002/cpt.910

Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., & Leeder, J. S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genetics in Medicine*, *19*(1), 69-76. https://doi.org/10.1038/gim.2016.80

Gaedigk, A., Whirl-Carrillo, M., Pratt, V. M., Miller, N. A., & Klein, T. E. (2020). PharmVar and the Landscape of Pharmacogenetic Resources. *Clin Pharmacol Ther*, *107*(1), 43-46. https://doi.org/10.1002/cpt.1654

Gamage, N., Barnett, A., Hempel, N., Duggleby, R. G., Windmill, K. F., Martin, J. L., & McManus, M. E. (2006). Human Sulfotransferases and Their Role in Chemical Metabolism. *Toxicological Sciences*, *90*(1), 5-22. https://doi.org/10.1093/toxsci/kfj061

Gardiner, S. J., & Begg, E. J. (2006). Pharmacogenetics, Drug-Metabolizing Enzymes, and Clinical Practice. *Pharmacological Reviews*, *58*(3), 521-590. https://doi.org/10.1124/pr.58.3.6

Georgitsi, M., Zukic, B., Pavlovic, S., & Patrinos, G. P. (2011). Transcriptional regulation and pharmacogenomics. *Pharmacogenomics*, *12*(5), 655-673. https://doi.org/10.2217/pgs.10.215

Gong, X., Liu, Y., Zhang, X., Wei, Z., Huo, R., Shen, L., He, L., & Qin, S. (2013). Systematic functional study of cytochrome P450 2D6 promoter polymorphisms in the Chinese Han population. *PLOS ONE*, *8*(2), e57764. https://doi.org/10.1371/journal.pone.0057764

Gough, A. C., Smith, C. A. D., Howell, S. M., Wolf, C. R., Bryant, S. P., & Spurr, N. K. (1993). Localization of the CYP2D Gene Locus to Human Chromosome 22q13.1 by Polymerase Chain Reaction, in

Situ Hybridization, and Linkage Analysis. *Genomics*, *15*(2), 430-432. https://doi.org/10.1006/geno.1993.1082

Gresham, D., Morar, B., Underhill, P. A., Passarino, G., Lin, A. A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P. J., Shen, P., Tournev, I., de Pablo, R., Kučinskas, V., Perez-Lezaun, A., Marushiakova, E., Popov, V., & Kalaydjieva, L. (2001). Origins and divergence of the Roma (gypsies). *Am J Hum Genet*, *69*(6), 1314-1331. https://doi.org/10.1086/324681

Guidice, J. M., Marez, D., Sabbagh, N., Legrand-Andreoletti, M., Spire, C., Alcaïde, E., Lafitte, J. J., & Broly, F. (1997). Evidence for CYP2D6 Expression in Human Lung. *Biochemical and Biophysical Research Communications*, *241*(1), 79-85. https://doi.org/10.1006/bbrc.1997.7775

Gusmao, A., Valente, C., Gomes, V., Alves, C., Amorim, A., Prata, M. J., & Gusmao, L. (2010). A genetic historical sketch of European Gypsies: The perspective from autosomal markers. *Am J Phys Anthropol*, *141*(4), 507-514. https://doi.org/10.1002/ajpa.21166

Hakkola, J., Pasanen, M., Hukkanen, J., Pelkonen, O., Mäenpää, J., Edwards, R. J., Boobis, A. R., & Raunio, H. (1996). Expression of xenobiotic-metabolizing cytochrome P450 Forms in human full-term placenta. *Biochemical Pharmacology*, *51*(4), 403-411. https://doi.org/10.1016/0006-2952(95)02184-1

Heim, M. H., & Meyer, U. A. (1992). Evolution of a highly polymorphic human cytochrome P450 gene cluster: CYP2D6. *Genomics*, *14*(1), 49-58. https://doi.org/10.1016/s0888-7543(05)80282-4

Hicks, J. K., Bishop, J. R., Sangkuhl, K., Müller, D. J., Ji, Y., Leckband, S. G., Leeder, J. S., Graham, R. L., Chiulli, D. L., LLerena, A., Skaar, T. C., Scott, S. A., Stingl, J. C., Klein, T. E., Caudle, K. E., Gaedigk, A., & Clinical Pharmacogenetics Implementation Consortium (2015). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clinical Pharmacology & Therapeutics*, *98*(2), 127-134. https://doi.org/10.1002/cpt.147

Hoffman, S. M. G., & Hu, S. (2007). Dynamic evolution of the CYP2ABFGST gene cluster in primates. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *616*(1-2), 133-138. https://doi.org/10.1016/j.mrfmmm.2006.11.004

Hrvatić, N., & Ivančić, S. (2000). Povijesno–socijalna obilježja Roma u Hrvatskoj. *Društvena istraživanja: časopis za opća društvena pitanja*, *9*(2-3 (46-47)), 251-266.

Hu, D. G., Marri, S., McKinnon, R. A., Mackenzie, P. I., & Meech, R. (2019). Deregulation of the Genes that Are Involved in Drug Absorption, Distribution, Metabolism, and Excretion in Hepatocellular Carcinoma. *Journal of Pharmacology and Experimental Therapeutics*, *368*(3), 363-381. https://doi.org/10.1124/jpet.118.255018

Hu, D. G., Meech, R., McKinnon, R. A., & Mackenzie, P. I. (2014). Transcriptional regulation of human UDP-glucuronosyltransferase genes. *Drug Metab Rev*, *46*(4), 421-458. https://doi.org/10.3109/03602532.2014.973037

Hu, S., Wang, H., Knisely, A. A., Reddy, S., Kovacevic, D., Liu, Z., & Hoffman, S. M. G. (2007). Evolution of the CYP2ABFGST gene cluster in rat, and a fine-scale comparison among rodent and primate species. *Genetica*, *133*(2), 215-226. https://doi.org/10.1007/s10709-007-9206-x

Ingelman-Sundberg, M. (2004). Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci*, *25*(4), 193-200. https://doi.org/10.1016/j.tips.2004.02.007

Ingelman-Sundberg, M., Sim, S. C., Gomez, A., & Rodriguez-Antona, C. (2007). Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacology & Therapeutics*, *116*(3), 496-526. https://doi.org/10.1016/j.pharmthera.2007.09.004

Jancova, P., Anzenbacher, P., & Anzenbacherova, E. (2010). Phase Ii Drug Metabolizing Enzymes. *Biomedical Papers*, *154*(2), 103-116. https://doi.org/10.5507/bp.2010.017

Kalaydjieva, L., Calafell, F., Jobling, M. A., Angelicheva, D., de Knijff, P., Rosser, Z. H., Hurles, M. E., Underhill, P., Tournev, I., Marushiakova, E., & Popov, V. (2001). Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet*, *9*(2), 97-104. https://doi.org/10.1038/sj.ejhg.5200597

Kalaydjieva, L., Gresham, D., & Calafell, F. (2001). Genetic studies of the Roma (Gypsies): a review. *BMC Medical Genetics*, *2*(1). https://doi.org/10.1186/1471-2350-2-5

Kalaydjieva, L., Morar, B., Chaix, R., & Tang, H. (2005). A newly discovered founder population: the Roma/Gypsies. *Bioessays*, *27*(10), 1084-1094. https://doi.org/10.1002/bies.20287

Kalow, W., & Bertilsson, L. (1994). Interethnic factors affecting drug response. *Advances in Drug Research*, *25*, 1-53.

Kane, M. (2021). CYP2D6 Overview: Allele and Phenotype Frequencies. In M. G. S. [Internet]. (Ed.), *Medical Genetics Summaries*. https://www.ncbi.nlm.nih.gov/books/NBK574601/

Kawashima, A., & Satta, Y. (2014). Substrate-dependent evolution of cytochrome P450: rapid turnover of the detoxification-type and conservation of the biosynthesis-type. *PLOS ONE*, *9*(6), e100059. https://doi.org/10.1371/journal.pone.0100059

Kimura, S., Umeno, M., Skoda, R. C., Meyer, U. A., & Gonzalez, F. J. (1989). The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *Am J Hum Genet*, *45*(6), 889-904. https://www.ncbi.nlm.nih.gov/pubmed/2574001

Klaric, I. M., Salihovic, M. P., Lauc, L. B., Zhivotovsky, L. A., Rootsi, S., & Janicijevic, B. (2009). Dissecting the molecular architecture and origin of Bayash Romani patrilineages: genetic influences from South-Asia and the Balkans. *Am J Phys Anthropol*, *138*(3), 333-342. https://doi.org/10.1002/ajpa.20933

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Arch Biochem Biophys*, *75*(2), 376-386. https://doi.org/10.1016/0003-9861(58)90436-3

Li, J., Lao, X., Zhang, C., Tian, L., Lu, D., & Xu, S. (2014). Increased genetic diversity of ADME genes in African Americans compared with their putative ancestral source populations and implications for pharmacogenomics. *BMC Genet*, *15*, 52. https://doi.org/10.1186/1471-2156-15-52

Li, J., Lou, H., Yang, X., Lu, D., Li, S., Jin, L., Pan, X., Yang, W., Song, M., Mamatyusupu, D., & Xu, S. (2014). Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *J Med Genet*, *51*(9), 614-622. https://doi.org/10.1136/jmedgenet-2014-102530

Li, J., Zhang, L., Zhou, H., Stoneking, M., & Tang, K. (2011). Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum Mol Genet*, *20*(3), 528-540. https://doi.org/10.1093/hmg/ddq498

Li, J., Zhang, L., Zhou, H., Stoneking, M., & Tang, K. (2011). Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum Mol Genet*, *20*(3), 528-540. https://doi.org/10.1093/hmg/ddq498

Liégeois, J.-P. (2009). *Romi u Europi*. Ibis.

Liégeois, J. (1989). Ciganos e Itinerantes: Dados socioculturais. *Dados sociopolíticos. Santa Casa da Misericórdia de Lisboa*.

López-García, M. A., Feria-Romero, I. A., Serrano, H., Rayo-Mares, D., Fagiolino, P., Vázquez, M., Escamilla-Núñez, C., Grijalva, I., Escalante-Santiago, D., & Orozco-Suarez, S. (2017). Influence of genetic variants of CYP2D6, CYP2C9, CYP2C19 and CYP3A4 on antiepileptic drug metabolism in pediatric patients with refractory epilepsy. *Pharmacological Reports*, *69*(3), 504-511. https://doi.org/10.1016/j.pharep.2017.01.007

Luizon, M. R., & Ahituv, N. (2015). Uncovering drug-responsive regulatory elements. *Pharmacogenomics*, *16*(16), 1829-1841. https://doi.org/10.2217/pgs.15.121

Lymperopoulos, A., McCrink, K. A., & Brill, A. (2015). Impact of CYP2D6 Genetic Variation on the Response of the Cardiovascular Patient to Carvedilol and Metoprolol. *Curr Drug Metab*, *17*(1), 30-36. https://doi.org/10.2174/1389200217666151105125425

Ma, Q., & Lu, A. Y. (2011). Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacol Rev*, *63*(2), 437-459. https://doi.org/10.1124/pr.110.003533

Madani, S., Paine, M. F., Lewis, L., Thummel, K. E., & Shen, D. D. (1999). Comparison of CYP2D6 content and metoprolol oxidation between microsomes isolated from human livers and small

intestines. *Pharmaceutical Research*, *16*(8), 1199-1205. https://doi.org/10.1023/a:1018989211864

Magalon, H., Patin, E., Austerlitz, F., Hegay, T., Aldashev, A., Quintana-Murci, L., & Heyer, E. (2008). Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet*, *16*(2), 243-251. https://doi.org/10.1038/sj.ejhg.5201963

Maisano Delser, P., & Fuselli, S. (2013). Human loci involved in drug biotransformation: worldwide genetic variation, population structure, and pharmacogenetic implications. *Hum Genet*, *132*(5), 563-577. https://doi.org/10.1007/s00439-013-1268-5

Malyarchuk, B. A., Grzybowski, T., Derenko, M. V., Czarny, J., & Miscicka-Sliwka, D. (2006). Mitochondrial DNA diversity in the Polish Roma. *Ann Hum Genet*, *70*(Pt 2), 195-206. https://doi.org/10.1111/j.1529-8817.2005.00222.x

Marushiakova, E., & Popov, V. (2001). Historical and ethnographic background: Gypsies, Roma, Sinti. *Between Past and Future: The Roma of Central and Eastern Europe*, 33.

Masimirembwa, C., Persson, I., Bertilsson, L., Hasler, J., & Ingelman-Sundberg, M. (1996). A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a black African population: association with diminished debrisoquine hydroxylase activity. *Br J Clin Pharmacol*, *42*(6), 713-719. https://doi.org/10.1046/j.1365-2125.1996.00489.x

Mendizabal, I., Valente, C., Gusmão, A., Alves, C., Gomes, V., Goios, A., Parson, W., Calafell, F., Alvarez, L., Amorim, A., Gusmão, L., Comas, D., & Prata, M. J. (2011). Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. *PLOS ONE*, *6*(1), e15988. https://doi.org/10.1371/journal.pone.0015988

Miklosich, F. (1873). *Über die Mundarten und die Wanderungen der Zigeuner Europa's: Beiträge zur Grammatik und zum Lexicon der Zigeunermundarten; 1. Beiträge zur Grammatik der Zigeunermundarten* (Vol. 2). Gerold.

Miksys, S., Rao, Y., Hoffmann, E., Mash, D. C., & Tyndale, R. F. (2002). Regional and cellular expression of CYP2D6 in human brain: higher levels in alcoholics. *Journal of Neurochemistry*, *82*(6), 1376-1387. https://doi.org/10.1046/j.1471-4159.2002.01069.x

Moorjani, P., Patterson, N., Loh, P. R., Lipson, M., Kisfali, P., Melegh, B. I., Bonin, M., Kádaši, L., Rieß, O., Berger, B., Reich, D., & Melegh, B. (2013). Reconstructing Roma History from Genome-Wide Data. *PLOS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0058633

Morar, B., Gresham, D., Angelicheva, D., Tournev, I., Gooding, R., Guergueltcheva, V., Schmidt, C., Abicht, A., Lochmuller, H., Tordai, A., Kalmar, L., Nagy, M., Karcagi, V., Jeanpierre, M., Herczegfalvi, A., Beeson, D., Venkataraman, V., Warwick Carter, K., Reeve, J., de Pablo, R., Kicinskas, V., Kalaydjieva, L. (2004). Mutation history of the roma/gypsies. *Am J Hum Genet*, *75*(4), 596-609. https://doi.org/10.1086/424759

Moya, G., Dorado, P., Ferreiro, V., Naranjo, M. E. G., Peñas-Lledó, E. M., & Llerena, A. (2016). High frequency of CYP2D6 ultrarapid metabolizer genotypes in an Ashkenazi Jewish population from Argentina. *The Pharmacogenomics Journal*, *17*(4), 378-381. https://doi.org/10.1038/tpj.2016.27

Nagaraj, S. H., & Toombs, M. (2021). The Gene-Drug Duality: Exploring the Pharmacogenomics of Indigenous Populations. *Front Genet*, *12*, 687116. https://doi.org/10.3389/fgene.2021.687116

Nagy, A., Sipeky, C., Szalai, R., Melegh, B. I., Matyas, P., Ganczer, A., Toth, K., & Melegh, B. (2015). Marked differences in frequencies of statin therapy relevant SLCO1B1 variants and haplotypes between Roma and Hungarian populations. *BMC Genet*, *16*, 108. https://doi.org/10.1186/s12863-015-0262-4

Naveen, A. T., Adithan, C., Soya, S. S., Gerard, N., & Krishnamoorthy, R. (2006). CYP2D6 Genetic Polymorphism in South Indian Populations. *Biological and Pharmaceutical Bulletin*, *29*(8), 1655-1658. https://doi.org/10.1248/bpb.29.1655

Nebert, D. W. (1997). Pharmacogenetics: 65 candles on the cake. *Pharmacogenetics*, *7*(6), 435-440.

Nebert, D. W. (1999). Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet*, *56*(4), 247-258. https://doi.org/10.1034/j.1399-0004.1999.560401.x

Nebert, D. W. (2002). Proposal for an allele nomenclature system based on the evolutionary divergence of haplotypes. *Hum Mutat*, *20*(6), 463-472. https://doi.org/10.1002/humu.10143

Nelson, D. R. (1999). Cytochrome P450 and the individuality of species. *Arch Biochem Biophys*, *369*(1), 1-10. https://doi.org/10.1006/abbi.1999.1352

Nelson, D. R. (2003). Comparison of P450s from human and fugu: 420 million years of vertebrate P450 evolution. *Archives of Biochemistry and Biophysics*, *409*(1), 18-24. https://doi.org/10.1016/s0003-9861(02)00553-2

Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., & Nebert, D. W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, *6*(1), 1-42. https://doi.org/10.1097/00008571-199602000-00002

Nelson, D. R., Zeldin, D. C., Hoffman, S. M. G., Maltais, L. J., Wain, H. M., & Nebert, D. W. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics*, *14*(1), 1-18. https://doi.org/10.1097/00008571-200401000-00001

Neve, E. P. A., & Ingelman-Sundberg, M. (2008). Intracellular transport and localization of microsomal cytochrome P450. *Anal Bioanal Chem*, *392*(6), 1075-1084. https://doi.org/10.1007/s00216-008-2200-z

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, *8*(11), 857-868. https://doi.org/10.1038/nrg2187

Nigam, S. K. (2014). What do drug transporters really do? *Nature Reviews Drug Discovery*, *14*(1), 29-44. https://doi.org/10.1038/nrd4461

Nishimura, M., Yaguti, H., Yoshitsugu, H., Naito, S., & Satoh, T. (2003). Tissue Distribution of mRNA Expression of Human Cytochrome P450 Isoforms Assessedby High-Sensitivity Real-Time Reverse Transcription PCR. *Yakugaku Zasshi*, *123*(5), 369-375. https://doi.org/10.1248/yakushi.123.369

Omura, T., & Sato, R. (1964). The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature. *J Biol Chem*, *239*, 2370-2378. https://www.ncbi.nlm.nih.gov/pubmed/14209971

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499-502. https://doi.org/10.1038/nature05295

Petrović, J., Pešić, V., & Lauschke, V. M. (2019). Frequencies of clinically important CYP2C19 and CYP2D6 alleles are graded across Europe. *European Journal of Human Genetics*, *28*(1), 88-94. https://doi.org/10.1038/s41431-019-0480-8

Pratt, V. M., Cavallari, L. H., Del Tredici, A. L., Gaedigk, A., Hachad, H., Ji, Y., Kalman, L. V., Ly, R. C., Moyer, A. M., Scott, S. A., van Schaik, R. H. N., Whirl-Carrillo, M., & Weck, K. E. (2021). Recommendations for Clinical CYP2D6 Genotyping Allele Selection. *The Journal of Molecular Diagnostics*, *23*(9), 1047-1064. https://doi.org/10.1016/j.jmoldx.2021.05.013

Prohaska, A., Racimo, F., Schork, A. J., Sikora, M., Stern, A. J., Ilardo, M., Allentoft, M. E., Folkersen, L., Buil, A., Moreno-Mayar, J. V., Korneliussen, T., Geschwind, D., Ingason, A., Werge, T., Nielsen, R., & Willerslev, E. (2019). Human Disease Variation in the Light of Population Genomics. *Cell*, *177*(1), 115-131. https://doi.org/10.1016/j.cell.2019.01.052

Prueksaritanont, T., Dwyer, L. M., & Cribb, A. E. (1995). (+)-Bufuralol 1'-hydroxylation activity in human and rhesus monkey intestine and liver. *Biochemical Pharmacology*, *50*(9), 1521-1525. https://doi.org/10.1016/0006-2952(95)02052-7

Radosavljevic, P. (2010). *The Language of the Boyash (Bayash) Roma on the Territory of the Republic of Croatia* University of Zagreb]. Croatia.

Rai, N., Chaubey, G., Tamang, R., Pathak, A. K., Singh, V. K., Karmin, M., Singh, M., Rani, D. S., Anugula, S., Yadav, B. K., Singh, A., Srinivasagan, R., Yadav, A., Kashyap, M., Narvariya, S., Reddy, A. G., van Driem, G., Underhill, P. A., Villems, R., Kivisild, T., Singh, L., Thangaraj, K. (2012). The Phylogeography of Y-Chromosome Haplogroup H1a1a-M82 Reveals the Likely Indian Origin of the European Romani Populations. *PLOS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048477

Raimundo, S., Fischer, J., Eichelbaum, M., Griese, E. U., Schwab, M., & Zanger, U. M. (2000). Elucidation of the genetic basis of the common 'intermediate metabolizer' phenotype for drug oxidation by CYP2D6. *Pharmacogenetics*, *10*(7), 577-581. https://doi.org/10.1097/00008571-200010000-00001

Rang, H. P. (2006). The receptor concept: pharmacology's big idea. *British Journal of Pharmacology*, *147*(S1), S9-S16. https://doi.org/10.1038/sj.bjp.0706457

Rees, D. C., Johnson, E., & Lewinson, O. (2009). ABC transporters: the power to change. *Nature Reviews Molecular Cell Biology*, *10*(3), 218-227. https://doi.org/10.1038/nrm2646

Reyniers, A. (1995). Gypsy populations and their movements within central and eastern Europe and towards some OECD countries.

Robarge, J. D., Li, L., Desta, Z., Nguyen, A., & Flockhart, D. A. (2007). The Star-Allele Nomenclature: Retooling for Translational Genomics. *Clinical Pharmacology & Therapeutics*, *82*(3), 244-248. https://doi.org/10.1038/sj.clpt.6100284

Robert, J., Le Morvan, V., Giovannetti, E., & Peters, G. J. (2014). On the use of pharmacogenetics in cancer treatment and clinical trials. *European Journal of Cancer*, *50*(15), 2532-2543. https://doi.org/10.1016/j.ejca.2014.07.013

Roses, A. D. (2000). Pharmacogenetics and the practice of medicine. *Nature*, *405*(6788), 857-865. https://doi.org/10.1038/35015728

Runcharoen, C., Fukunaga, K., Sensorn, I., Iemwimangsa, N., Klumsathian, S., Tong, H., Vo, N. S., Le, L., Hlaing, T. M., Thant, M., Zain, S. M., Mohamed, Z., Pung, Y. F., Capule, F., Nevado, J., Jr, Silao, C. L., Al-Mahayri, Z. N., Ali, B. R., Yuliwulandari, R., Prayuni, K., Zahroh, H., Noor, D. Z. A., Xangsayarath, P., Xayavong, D., Kounnavong, S., Sayasone, S., Kordou, Z., Liopetas, T., Tsikrika, A., Tsermpini, E., Korominam M., Mitropoulou, C., Patrinos, G. P., Kesornsit, A., Charoenyingwattana, A., Wattanapokayakit, S., Mahasirimongkol, S., Mushiroda, T., Chantratita, W. (2021). Prevalence of pharmacogenomic variants in 100 pharmacogenes among Southeast Asian populations under the collaboration of the Southeast Asian Pharmacogenomics Research Network (SEAPharm). *Human Genome Variation*, *8*(1). https://doi.org/10.1038/s41439-021-00135-z

Salihovic, M. P., Baresic, A., Klaric, I. M., Cukrov, S., Lauc, L. B., & Janicijevic, B. (2011). The role of the Vlax Roma in shaping the European Romani maternal genetic history. *Am J Phys Anthropol*, *146*(2), 262-270. https://doi.org/10.1002/ajpa.21566

Siegle, I., Fritz, P., Eckhardt, K., Zanger, U. M., & Eichelbaum, M. (2001). Cellular localization and regional distribution of CYP2D6 mRNA and protein expression in human brain. *Pharmacogenetics*, *11*(3), 237-245. https://doi.org/10.1097/00008571-200104000-00007

Sindrup, S. H., Poulsen, L., Brosen, K., Arendt-Nielsen, L., & Gram, L. F. (1993). Are poor metabolisers of sparteine/debrisoquine less pain tolerant than extensive metabolisers? *Pain*, *53*(3), 335-339. https://doi.org/10.1016/0304-3959(93)90229-i

Sipeky, C., Csongei, V., Jaromi, L., Safrany, E., Maasz, A., Takacs, I., Beres, J., Fodor, L., Szabo, M., & Melegh, B (2011). Genetic variability and haplotype profile of MDR1 (ABCB1) in Roma and Hungarian population samples with a review of the literature. *Drug Metab Pharmacokinet*, *26*(2), 206-215.

Sipeky, C., Csongei, V., Jaromi, L., Safrany, E., Polgar, N., Lakner, L., Szabo, M., Takacs, I., & Melegh, B. (2009). Vitamin K epoxide reductase complex 1 (VKORC1) haplotypes in healthy Hungarian

and Roma population samples. *Pharmacogenomics*, *10*(6), 1025-1032. https://doi.org/10.2217/pgs.09.46

Sipeky, C., Weber, A., Szabo, M., Melegh, B. I., Janicsek, I., Tarlos, G., Szabo, I., Sumegi, K., & Melegh, B. (2013). High prevalence of CYP2C19*2 allele in Roma samples: study on Roma and Hungarian population samples with review of the literature. *Mol Biol Rep*, *40*(8), 4727-4735. https://doi.org/10.1007/s11033-013-2569-4

Sistonen, J., Sajantila, A., Lao, O., Corander, J., Barbujani, G., & Fuselli, S. (2007). CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenet Genomics*, *17*(2), 93-101. https://doi.org/10.1097/01.fpc.0000239974.69464.f2

Skarić-Jurić, T., Klarić, I. M., Narancić, N. S., Drmić, S., Salihović, M. P., Lauc, L. B., Milicić, J., Barabalić, M., Zajc, M., & Janićijević, B. (2007). Trapped between tradition and transition--anthropological and epidemiological cross-sectional study of Bayash Roma in Croatia. *Croat Med J*, *48*(5), 708-719.

Škarić-Jurić, T., Tomas, Ž., Zajc Petranović, M., Božina, N., Smolej Narančić, N., Janićijević, B., & Salihović, M. P. (2018). Characterization of ADME genes variation in Roma and 20 populations worldwide. *PLOS ONE*, *13*(11), e0207671. https://doi.org/10.1371/journal.pone.0207671

Smale, S. T., & Kadonaga, J. T. (2003). The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry*, *72*(1), 449-479. https://doi.org/10.1146/annurev.biochem.72.121801.161520

Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, *73*(5), 1162-1169. https://doi.org/10.1086/379378

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, *68*(4), 978-989. https://doi.org/10.1086/319501

Stingl, J. C., Brockmöller, J., & Viviani, R. (2012). Genetic variability of drug-metabolizing enzymes: the dual impact on psychiatric therapy and regulation of brain function. *Molecular Psychiatry*, *18*(3), 273-287. https://doi.org/10.1038/mp.2012.42

Stojanovic Markovic, A., Zajc Petranovic, M., Tomas, Z., Puljko, B., Setinc, M., Skaric-Juric, T., & Pericic Salihovic, M. (2022). Untangling SNP Variations within CYP2D6 Gene in Croatian Roma. *J Pers Med*, *12*(3). https://doi.org/10.3390/jpm12030374

Stojanović Marković, A., Zajc Petranović, M., Škarić-Jurić, T., Celinšćak, Ž., Šetinc, M., Tomas, Ž., & Peričić Salihović, M. (2022). Relevance of CYP2D6 gene variants in population genetic differentiation. *Pharmaceutics*, *14*(11). https://doi.org/10.3390/pharmaceutics14112481

Stojanović Marković, A., Zajc Petranović, M., Škobalj, M., Poloni, E. S., Pichler Oberški, L., Škarić-Jurić, T., & Peričić Salihović, M. (2022). From dietary adaptation in the past to drug metabolism of today: An example of NAT genes in the Croatian Roma. *American Journal of Biological Anthropology*, *178*(1), 140-153. https://doi.org/10.1002/ajpa.24483

Suarez-Kurtz, G., Paula, D. P., & Struchiner, C. J. (2014). Pharmacogenomic implications of population admixture: Brazil as a model case. *Pharmacogenomics*, *15*(2), 209-219. https://doi.org/10.2217/pgs.13.238

Teixeira, J., Amorim, A., J. Prata, M., & Quental, S. (2015). Pharmacogenetic Polymorphisms in a Portuguese Gypsy Population. *Current Pharmacogenomics and Personalized Medicine*, *13*(1), 36-40. https://doi.org/10.2174/1875692113666150703180101

Tomas, Z., Kuhanec, A., Skaric-Juric, T., Petranovic, M. Z., Narancic, N. S., Janicijevic, B., & Salihovic, M. P. (2017). Distinctiveness of the Roma population within CYP2B6 worldwide variation. *Pharmacogenomics*, *18*(17), 1575-1587. https://doi.org/10.2217/pgs-2017-0105

Touw, D. J., Neef, C., Thomson, A. H., & Vinks, A. A. (2005). Cost-Effectiveness of Therapeutic Drug Monitoring. *Therapeutic Drug Monitoring*, *27*(1), 10-17. https://doi.org/10.1097/00007691-200502000-00004

Urban, P., Lautier, T., Pompon, D., & Truan, G. (2018). Ligand Access Channels in Cytochrome P450 Enzymes: A Review. *International Journal of Molecular Sciences*, *19*(6). https://doi.org/10.3390/ijms19061617

Visel, A., Akiyama, J. A., Shoukry, M., Afzal, V., Rubin, E. M., & Pennacchio, L. A. (2009). Functional autonomy of distant-acting human enhancers. *Genomics*, *93*(6), 509-513. https://doi.org/10.1016/j.ygeno.2009.02.002

Vogel, F. (1959). Moderne Probleme der Humangenetik. In *Ergebnisse der Inneren Medizin und Kinderheilkunde* (pp. 52-125). https://doi.org/10.1007/978-3-642-94744-5_2

Wang, B., Yang, L.-P., Zhang, X.-Z., Huang, S.-Q., Bartlam, M., & Zhou, S.-F. (2009). New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. *Drug Metab Rev*, *41*(4), 573-643. https://doi.org/10.1080/03602530903118729

Wang, D., Poi, M. J., Sun, X., Gaedigk, A., Leeder, J. S., & Sadee, W. (2014). Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum Mol Genet*, *23*(1), 268-278. https://doi.org/10.1093/hmg/ddt417

Weber, A., Szalai, R., Sipeky, C., Magyari, L., Melegh, M., Jaromi, L., Matyas, P., Duga, B., Kovesdi, E., Hadzsiev, K., & Melegh, B. (2015). Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples. *Pharmacological Reports*, *67*(3), 460-464. https://doi.org/10.1016/j.pharep.2014.11.006

Wen, Y. F., Gaedigk, A., Boone, E. C., Wang, W. Y., & Straka, R. J. (2022). The Identification of Novel CYP2D6 Variants in US Hmong: Results From Genome Sequencing and Clinical Genotyping. *Front Pharmacol*, *13*, 867331. https://doi.org/10.3389/fphar.2022.867331

Williams, I. S., Gatchie, L., Bharate, S. B., & Chaudhuri, B. (2018). Biotransformation, Using Recombinant CYP450-Expressing Baker's Yeast Cells, Identifies a Novel CYP2D6.10A122V Variant Which Is a Superior Metabolizer of Codeine to Morphine Than the Wild-Type Enzyme. *ACS Omega*, *3*(8), 8903-8912. https://doi.org/10.1021/acsomega.8b00809

Woodland, C., Huang, T. T., Gryz, E., Bendayan, R., & Fawcett, J. P. (2008). Expression, Activity and Regulation of CYP3A in Human and Rodent Brain. *Drug Metab Rev*, *40*(1), 149-168. https://doi.org/10.1080/03602530701836712

Yasukochi, Y., & Satta, Y. (2011). Evolution of the CYP2D gene cluster in humans and four non-human primates. *Genes & Genetic Systems*, *86*(2), 109-116. https://doi.org/10.1266/ggs.86.109

Zajc Petranovic, M., Tomas, Z., Skaric Juric, T., Smolej Narancic, N., Janicijevic, B., & Pericic Salihovic, M. (2018). The variation of CYP2C19 gene in the Roma population from Croatia. *Molecular and experimental biology in medicine*, *1*(2), 32-37.

Zajc Petranovic, M., Tomas, Z., Skaric Juric, T., Smolej Narancic, N., Janicijevic, B., Stojanovic Markovic, A., & Pericic Salihovic, M. (2019). The variability of multi-drug resistance ABCB1 gene in the Roma population from Croatia. *Molecular and experimental biology in medicine*, *2*(1), 10-18.

Zajc Petranović, M., Rizzieri, A. E., Sivaraj, D., Smolej Narančić, N., Škarić-Jurić, T., Celinšćak, Ž., Stojanović Marković, A., Peričić Salihović, M., Kalászi, J., Kalászi, M., Lin, J. Q., Mehta, S., Burleson, J., & Rizzieri, D. A. (2021). CVD Risk Factors in the Ukrainian Roma and Meta-Analysis of Their Prevalence in Roma Populations Worldwide. *Journal of Personalized Medicine*, *11*(11). https://doi.org/10.3390/jpm11111138

Zanger, U. M., Fischer, J., Raimundo, S., Stuven, T., Evert, B. O., Schwab, M., & Eichelbaum, M. (2001). Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics*, *11*(7), 573-585. https://doi.org/10.1097/00008571-200110000-00004

Zanger, U. M., Raimundo, S., & Eichelbaum, M. (2004). Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *369*(1), 23-37. https://doi.org/10.1007/s00210-003-0832-2

Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther*, *138*(1), 103-141. https://doi.org/10.1016/j.pharmthera.2012.12.007

Zanger, U. M., Turpeinen, M., Klein, K., & Schwab, M. (2008). Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal Bioanal Chem*, *392*(6), 1093-1108. https://doi.org/10.1007/s00216-008-2291-6

Zhang, H., De, T., Zhong, Y., & Perera, M. A. (2019). The Advantages and Challenges of Diversity in Pharmacogenomics: Can Minority Populations Bring Us Closer to Implementation? *Clinical Pharmacology & Therapeutics*, *106*(2), 338-349. https://doi.org/10.1002/cpt.1491

Zhao, M., Ma, J., Li, M., Zhang, Y., Jiang, B., Zhao, X., Huai, C., Shen, L., Zhang, N., He, L., & Qin, S. (2021). Cytochrome P450 Enzymes and Drug Metabolism in Humans. *International Journal of Molecular Sciences*, *22*(23). https://doi.org/10.3390/ijms222312808

Zhou, S. (2018). *Cytochrome P450 2D6: Structure, Function, Regulation and Polymorphism*. CRC Press. https://doi.org/10.1201/9781315372983

Zhou, Y., & Lauschke, V. M. (2018). Comprehensive overview of the pharmacogenetic diversity in Ashkenazi Jews. *J Med Genet*, *55*(9), 617-627. https://doi.org/10.1136/jmedgenet-2018-105429

6. CURRICULUM VITAE

Anita Stojanović Marković was born in Slavonski Brod in 1989, where she completed elementary school and Classical Gymnassium “fra Marijan Lanosović”. In 2015 she graduated from Faculty of Science, Department of Biology, University of Zagreb and obtained Master’s degree in Molecular Biology with the topic of her thesis “Association of gestational diabetes with serotonin transporter gene DNA methylation in placenta”. During her studies together with her colleagues, she was awarded the Rector's Award for the research project “Regulation of *MGAT3* gene expression in malignant tumours derived from epithelial cells” that was perfomed at the Division of Molecular Biology of the Faculty of Science under the mentorship of prof. Petra Korać, PhD. Anita Stoajnović Marković did a 6 month Eramsus+ internship in Vienna, at the Department of Neuroimmunology, Center for Brain Research, Medical University Vienna at the end of her graduate studies.

From February 2017 to August 2018, she worked as an analyst for food safety at the Laboratory for Molecular Diagnostics, Croatiakontrola d.o.o. Since September 2018, she has been employed at the Institute for Anthropological Research as an assistant/doctoral student under the supervision of prof. Marijana Peričić Salihović, PhD as part of the "Young researcher’s career development project – training new doctoral students”. In the same year, she enrolled in a postgraduate doctoral programme in Biology at the Department of Biology, Faculty of Science, University of Zagreb. During her postgraduate studies, for 3.5 months she trained in the field of computational population genetics at the Institute of Genomics, University of Tartu (Estonia), financed through an Erasmus+ scholarship. She was one of the mentors in practical courses in the course "Human Genome" led by prof. Marijana Peričić Salihović

Anita Stojanović Marković published 8 scientific papers in peer-reviewed journals, 4 of them as the first author. She participated in numerous workshops and scientific meetings. She is a member of scientific associations: European Anthropological Association and Pharmacogenomics Global Research Network.

7. PROŠIRENI SAŽETAK

Gen *CYP2D6* je jedan od najpolimorfnijih ADME gena odgovornih za apsorpciju, metabolizam, distribuciju i izlučivanje (ekskreciju) lijekova. Ovi geni kodiraju enzime metabolizma faze I i II te transportere lijekova i modifikatore. Među najvažnijim enzimima faze I za metabolizam lijekova su enzimi citokoroma P450 (CYP). Gen *CYP2D6* je jedini funkcionalni gen podobitelji CYP2D, a uz to je izuzetno polimorfan s više od 100 otkrivenih varijanti i podvarijanti te 170 haplotipova koji se prema farmakogenetičkoj nomenklaturi nazivaju zvjezdasti aleli. Zvjezdasti aleli se mogu s obzirom na enzimsku funkciju razvrstati u one koji rezultiraju povećanom, smanjenom, normalnom enzimskom aktivnošću ili s gubitkom funkcije (tzv. null aleli). Prema aktivnosti, enzim CYP2D6 se grupira u četiri različita metabolizatora lijekova (ultra brzi, normalni, srednji i slabi metabolizator) i dvije podklase (srednji do normalnog i spori do srednjeg). Ovaj gen kodira za istoimeni enzim odgovaran za metabolizam oko 25% klinički propisanih lijekova iako čini samo 2% svih citokroma u jetri čovjeka.

Prema bazi podataka GeneCards, postoji 87 lokusa u regijama promotora i pojačivača koje su vezane uz ekspresiju gena *CYP2D6*, no veza između aktivnosti promotora/pojačivača i varijacija u genu *CYP2D6* za metabolizam lijekova nije dovoljno proučena. Polimorfizmi jednog nukleotida (eng. SNP) iz regije pojačivača ili promotora mogu biti neravnotežno povezani (LD) sa SNP-ovima koji definiraju zvjezdaste alele iz regije gena *CYP2D6* što može utjecati na metabolizam.

Geni ADME su unatoč svojoj dokazanoj i važnoj funkcionalnoj ulozi slabo istraženi u izoliranim populacijama. Jedan od primjera takve populacije su Romi, transnacionalna manjinska populacija prisutna u mnogim zemljama svijeta, čije je podrijetlo iz Indije iz koje su u Europu došli oko 11. stoljeća, putem kroz centralnu Aziju i današnju Tursku. Procjenjuje se kako danas u svijetu živi oko 15 milijuna Roma, od kojih čak 12 milijuna u Europi. Romi u Hrvatskoj su jedna od 22 priznate nacionalne manjine. Prema popisu stanovništva iz 2021., u Hrvatskoj živi 17.980 Roma, no zbog raznih povijesnih razloga evidencija pripadnika romske populacije nije uvijek pouzdana - neslužbeno je u Hrvatskoj između 30.000 i 40.000 Roma. Romi u Hrvatskoj se prema lingvističkim obilježjima dijele u dvije skupine (Vlaški i Balkanski Romi). Vlaški Romi je naziv za skupine koje dijele povijest 500-godišnjeg ropstva u današnjoj Rumunjskoj, što je dovelo do njihova odvajanja u mnogo manjih podgrupa s različitim običajima i načinom života. Vlaški Romi govore arhaičnim staro rumunjskim jezikom, ljimba d'bjaš, koji dijelimo na više različitih dijalekta. Erdeljskim dijalektom se služe Romi iz Međimurja, a baranjskim mutenskim Romi iz Baranje. Drugu veliku skupinu

čine Balkanski Romi, čiji su preci naselili područje Balkana još u 11. stoljeću za vrijeme Otomanskog Carstva. Ovi Romi se u svakodnevnom govoru služe dijalektima romskog jezika romani chib.

Romske skupine imaju zajedničko porijeklo, ali ne i blisku prošlost radi razdvajanja seobenih putova kroz povijest, što je dovelo do razlika u genetičkoj strukturi skupina. U zalihi gena romskih populacija uočen je određeni stupanj miješanja s okolnim većinskim stanovništvom, iako su u najvećoj mjeri sačuvali svoju genetičku i socio-kulturnu izoliranost. Na razini cjelokupne romske populacije postoji relativna homogenost, ali postoji i jasno izražena različitost među grupama.

Cilj ove doktorske disertacije je bio ispitati utjecaj socio-kulturnih i migracijskih obilježja romske populacije Hrvatske na njihov farmakogenetski profil analizom gena *CYP2D6*. Analize su provedene na uzorcima DNA pripadnika romskih populacija Hrvatske koji pripadaju trima različitim migracijskim i dijalektalnim skupinama. Hrvatsku romsku populaciju čine dvije velike skupine: Balkanski Romi i dvije dijalektalne skupine Vlaških Roma.

Rezultati ovog istraživanja su pokazali utjecaj demografske povijesti, posebno migracija i endogamije, na distribuciju varijanti gena *CYP2D6* u tri socio-kulturno različite proučavane romske populacije. To se naročito vidi u nakupljanju globalno rijetkih varijanti, što je posljedica genskog pomaka. Učestalosti polimorfizama koje su slične azijskim populacijama, kao i povećana učestalost zvjezdastih alela **10* i **41*, ukazuju na njihovo južnoazijsko porijeklo. Fenotipski gledano, kod Roma je najučestaliji normalni metabolizator (51,6-65,1%), dok je spori najmanje zastupljen (0-7,4%). Sličnost Roma europskim i azijskim populacijama je vidljiva iz obrasca LD-a genske regije *CYP2D6*. Tri romske skupine se značajno razlikuju ($p<0,0001$) u distribuciji pet najzastupljenijih haplotipova (**1, *2, *4, *10, *41*), što potvrđuje potrebu da se u istraživanjima proučavaju kao zasebne grupe radi kvalitetnije upotrebe lijekova koje metabolizira *CYP2D6*.

Prema našim saznanjima ovo je prvo opsežno istraživanje gena *CYP2D6* kod Roma. Rezultati daju važan doprinos farmakogenetičkim spoznajama izoliranih populacija te potencijalno točniju primjenu lijekova u ovoj osjetljivoj populaciji.