

ПОЛИМОРФИЗМ 27 АУТОСОМНЫХ STR-ЛОКУСОВ У НАСЕЛЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ ПО ДАННЫМ МАССОВОГО ПАРАЛЛЕЛЬНОГО СЕКВЕНИРОВАНИЯ¹

© 2023 г. С. А. Котова¹, *, А. С. Парфёнова¹, Т. В. Забавская¹,
В. И. Рыбакова¹, Е. А. Спивак¹, С. А. Полевой², А. В. Луговнёв³

¹Научно-практический центр государственного комитета судебных экспертиз
Республики Беларусь, Минск, 220114 Республика Беларусь

²Масариков университет, Брно, 60177 Чехия

³Государственный комитет судебных экспертиз
Республики Беларусь, Минск, 220073 Республика Беларусь

*e-mail: svetlkotova@mail.ru

Поступила в редакцию 27.04.2022 г.

После доработки 26.05.2022 г.

Принята к публикации 23.06.2022 г.

С использованием технологии массового параллельного секвенирования (МПС) проведено исследование полиморфизма 27 аутосомных STR-локусов коммерческой панели ForenSeq DNA Signature Prep Kit у 733 неродственных индивидов, представляющих население Республики Беларусь, и сформирована популяционная база частот МПС-аллелей для вероятностных расчетов при судебно-экспертной идентификации личности и установлении биологического родства. Соответствие между генотипами, полученными МПС и капиллярным электрофорезом (КЭ), составило 99.96%. Для восьми локусов (*D12S391*, *D21S11*, *D2S1338*, *vWA*, *D3S1358*, *D8S1179*, *D13S317*, *D9S1122*) количество МПС-аллелей увеличилось более чем в 2 раза. В исследованной выборке идентифицировано 13 новых аллелей, отсутствующих в каталоге STRSeq BioProject международной онлайн-базы данных STRbase 2.0. Средняя вероятность случайного совпадения 27-локусных МПС-профилей уменьшилась с 1.43×10^{-31} до 2.89×10^{-35} , а комбинированный индекс родства возрос с 2.08×10^{10} до 3.25×10^{12} по сравнению с данными КЭ.

Ключевые слова: аутосомный STR-локус, полиморфизм, изометрические аллели, массовое параллельное секвенирование.

DOI: 10.31857/S0016675823030074, **EDN:** IONQNZ

Микросателлитные (short tandem repeat, STR) локусы аутосом являются одними из наиболее широко применяемых маркеров в экспертном ДНК-анализе. Благодаря большому числу аллелей на локус и кодоминантному наследованию они позволяют с высокой вероятностью идентифицировать личность, а также устанавливать биологическое родство [1]. Уже более чем два десятилетия стандартной технологией получения STR-профиля ДНК является капиллярный электрофорез (КЭ), позволяющий выявлять полиморфизм микросателлитных локусов, обеспечивающийся вариативностью числа тандемных повторов. Благодаря стремительному развитию и внедрению в практику методов массового параллельного секвенирования (МПС) стало возможно получать детальную ин-

формацию о нуклеотидной последовательности традиционных криминалистических STR-маркеров [2]. К основным преимуществам технологии МПС судебные генетики также относят возможность одновременного анализа большого количества микросателлитов с длинами аллелей до 400 пн, что в сочетании с высокой чувствительностью метода значительно повышает эффективность анализа деградированной ДНК, а выявление значительного количества изоаллелей, обусловленных различиями в области как тандемного повтора, так и фланкирующих регионов, способствует лучшему разделению смесей в случае маскировки аллелей при использовании капиллярного электрофореза. Это привело к возникновению различных МПС-платформ, использующих маркеры, подходящие для криминалистических целей. Примером двух таких платформ являются MiSeq FGx Forensic Genomics System (Illumina, США) и Ion GeneStu-

¹ Дополнительная информация для этой статьи доступна по doi 10.31857/S0016675823030074 для авторизованных пользователей.

dio S5 System (Thermo Fisher Scientific, США), которые предлагают криминалистам наборы ForenSeq DNA Signature Prep kit (Verogen, США) и Precision ID GlobalFiler™ NGS STR Panel v2 (Thermo Fisher Scientific) [3, 4]. Вышеуказанные компании сотрудничают с научным консорциумом для совершенствования своих продуктов и включают в наборы маркеры, наиболее релевантные по литературным данным. Благодаря этому для ForenSeq DNA Signature Prep Kit и Precision ID GlobalFiler™ NGS STR Panel v2 общими являются 22 аутосомных STR-маркера.

С 2018 г. методы МПС начали активно внедряться в экспертную практику в Республике Беларусь. Для того чтобы в полной мере использовать более высокий дискриминирующий потенциал, предоставляемый полученными с использованием МПС-технологий результатами исследования STR-локусов, необходимо обновление данных частот встречаемости, которые бы содержали сведения о секвенированных аллелях. Все ранее проведенные в Республике Беларусь популяционные исследования базировались на классических технологиях, которые таких сведений не содержат и таким образом не могут обеспечить необходимый уровень информативности [5, 6].

В настоящем исследовании представлены результаты массового параллельного секвенирования 27 аутосомных STR-локусов панели ForenSeq DNA Signature Prep Kit (Illumina) у 733 индивидов, проживающих в Республике Беларусь. Для верификации результатов секвенирования и обеспечения их совместимости с ранее разработанной базой данных [5, 6] генотипы 165 образцов были проверены на конкордантность путем генотипирования 22 перекрывающихся аутосомных локусов с использованием набора для капиллярного электрофореза PowerPlex Fusion 6C System (Promega Corporation, США). Все МПС-аллели, выявленные в белорусской популяции, сравнивали с опубликованными и включенными в международные онлайн-базы данными [7] с целью обнаружения не описанных ранее вариантов.

МАТЕРИАЛЫ И МЕТОДЫ

Образцы биологического материала и выделение ДНК

Образцы венозной крови отбирали у анонимных доноров после получения письменного информированного согласия на проведение исследования и опубликование его результатов в открытой печати с соблюдением полной конфиденциальности полученной от участников информации. Всего исследовано 733 неродственных индивида мужского пола, проживающих в шести регионах страны: юго-запад (Брестская область) — 137; юго-восток (Гомельская область) — 111; запад (Грод-

ненская область) — 122; восток (Могилевская область) — 118; север (Витебская область) — 100 и центральный регион (Минская область) — 145.

Забор венозной крови осуществлялся с помощью систем типа Vacutainer (Becton Dickinson International, США) при участии специалистов Республиканского научно-практического центра трансфузиологии и медицинских биотехнологий. До выделения ДНК все образцы крови хранили при температуре -20°C .

ДНК выделяли из лейкоцитарной фракции венозной крови по методу, описанному в [8], после предварительного этапа гемолиза эритроцитов. Очистку ДНК проводили с помощью органических растворителей [9]. Концентрацию ДНК измеряли на капельном спектрофотометре DeNovix (DeNovix, США) в соответствии с протоколом производителя.

Подготовка геномных библиотек

Подготовку библиотек для секвенирования аутосомных STR-маркеров осуществляли с использованием набора ForenSeq™ DNA Signature Prep Kit (Verogen, США) согласно инструкции производителя [10]. Предварительно измеряли концентрацию образцов ДНК на приборе Quantus™ Fluorometer (Promega, США). Образцы ДНК нормализовали до конечной концентрации 0.2 нг/мкл. Для первой ПЦР использовали пул праймеров *B*, предназначенный для совместной амплификации 231 маркера ядерной ДНК, включая 27 аутосомных STR, 24 Y-STR, 7 X-хромосомных STR, 94 идентифицирующих SNP, 56 геногеографических SNP, 22 фенотипических SNP и амелогенин. В качестве положительного контроля реакции амплификации использовали 1 нг контрольной ДНК 2800M (Promega Corporation, США), в качестве отрицательного — деионизированную воду. Амплификацию проводили на термоциклере C1000 Touch™ Thermal Cycler (Bio-Rad, США) согласно инструкции производителя [10]. Целевое обогащение, индексирование, очистку и нормализацию геномных библиотек выполняли следуя протоколу производителя. Готовые нормализованные библиотеки смешивали в одной пробирке, проводили денатурацию и загружали на проточную ячейку MiSeq FGx® Reagents Kit (Verogen). Для всех пулов число библиотек соответствовало рекомендованному производителем числу в 32 библиотеки. Секвенирование проводили в режиме экспертно-го ДНК-анализа.

Секвенирование и анализ результатов

Секвенирование библиотек проводили на платформе MiSeq FGx™ (Illumina Inc., США), обработку результатов выполняли с помощью интегрированного в систему программного обеспече-

ния ForenSeq Universal Analysis Software (ForenSeq UAS) (Illumina Inc.). Генотипы образцов по ауто-сомным STR-маркерам *D3S1358*, *vWA*, *D16S539*, *CSF1PO*, *TPOX*, *D8S1179*, *D21S11*, *D18S51*, *D2S441*, *D19S433*, *TH01*, *FGA*, *D22S1045*, *D5S818*, *D13S317*, *D7S820*, *D10S1248*, *D1S1656*, *D12S391*, *D2S1338*, *D6S1043*, *PentaD*, *PentaE*, *D4S2408*, *D9S1122*, *D17S1301*, *D20S482* анализировали на основе стандартных выходных данных, включавших отчет о генотипе на уровне проекта и отчет о деталях образца на уровне генотипа, сгенерированных ForenSeq UAS на листах Microsoft Excel. По умолчанию минимальные пороговые значения для анализа (АП) и интерпретации данных (ИП) принимали 10 и 30 прочтений соответственно для всех аутосомных STR-локусов.

Генотипы регистрировали как число tandemных повторов в вариабельной области микросателлита, автоматически зарегистрированное ForenSeq UAS; также анализировали первичную последовательность вариабельной области STR-маркера. Для локусов *D12S317*, *D18S51*, *D19S433*, *D5S818*, *D7S820* и *vWA* дополнительно изучали нуклеотидную последовательность доступного для анализа небольшого фланкирующего участка, непосредственно прилегающего к tandemной области. Однонуклеотидные варианты полиморфизма, обнаруженные в области tandemного повтора и на доступных для анализа фланкирующих участках микросателлитов, регистрировали как вариант замены по отношению к референсному геному GRCh37.p13 [11]. Последовательности аллельных вариантов описывали согласно требованиям Международного общества по судебной генетике [12]; аллелям присваивали численно-буквенные значения.

Верификация с данными капиллярного электрофореза

Для верификации полученных МПС-генотипов контрольную ДНК 2800М и 165 образцов ДНК популяционной выборки, отобранных случайным образом, генотипировали с использованием набора PowerPlex® Fusion 6C System (Promega Corporation), который включает 22 общих с ForenSeq™ DNA Signature Prep Kit аутосомных локуса: *D3S1358*, *vWA*, *D16S539*, *CSF1PO*, *TPOX*, *D8S1179*, *D21S11*, *D18S51*, *D2S441*, *D19S433*, *TH01*, *FGA*, *D22S1045*, *D5S818*, *D13S317*, *D7S820*, *D10S1248*, *D1S1656*, *D12S391*, *D2S1338*, *PentaD*, *PentaE*. Все эксперименты проводили в соответствии с протоколами производителя реагентов. Продукты ПЦР разделяли методом КЭ на генетическом анализаторе AB3500 (Thermo Fischer Scientific). Генотипы образцов определяли с использованием программного обеспечения GeneMapper ID-X 1.4 (Thermo Fisher Scientific). Результаты генотипирования, полученные методами МПС и КЭ, сравнивали,

дискордантность определяли как любой случай, когда аллель, обнаруженный с помощью одного метода, не обнаруживался с помощью другого. В случаях, когда глубина прочтения для одного из аллелей профиля была ниже ИП и аллель не регистрировался программным обеспечением, несоответствие считали обусловленным “выпадением” аллеля (*allele drop-out*) [13].

Популяционные данные и криминалистические параметры информативности

Изменение уровня генетического разнообразия и криминалистических параметров оценивали сравнивая значения наблюдаемой гетерозиготности (H_0) и вероятности случайного совпадения (MP), рассчитанные для двух массивов МПС-генотипов, включающих: 1) аллели, определенные по числу tandemных повторов (КЭ-аллели) и 2) аллели, определенные по нуклеотидной последовательности (МПС-аллели).

Частоты аллелей для 27 аутосомных STR-локусов и наблюдаемую гетерозиготность (H_0) рассчитывали на основе результатов МПС-генотипирования 733 популяционных образцов средстами программного пакета GenAlEx 6.5 [14].

Криминалистические параметры информационной значимости: силу различения двух генотипов (PD), силу исключения (PE), типичный индекс отцовства (TPI) и их комбинированные значения рассчитывали с использованием алгоритмов FORSTAT v1.0 [15].

Оценку уровня генетической подразделенности популяции (F_{ST}) проводили методом анализа молекулярной дисперсии (Analysis of Molecular Variance, AMOVA) в программном пакете Arlequin 3.5 [16].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

С целью формирования базы данных частот МПС-аллелей 27 криминалистических STR-локусов для резидентного населения Республики Беларусь в ходе настоящего исследования секвенировано 39 582 фрагмента ДНК для 733 индивидов и получены данные о 19 767 генотипах (99.87% всех возможных STR-генотипов), а также изучена внутренняя структура области tandemного повтора аллельных вариантов микросателлитов. Для 24 образцов данные о генотипах отсутствовали в локусах *D1S1656* (12 образцов), *PentaE* (7 образцов), *PentaD* (4 образца) и *D21S11* (1 образец). При анализе файлов отчетов, сгенерированных ForenSeq UAS, во всех случаях последовательности аллелей для этих образцов были обнаружены, однако по-скольку уровень прочтений для них был ниже установленного ИП (30 ридов) программным обеспечением они не регистрировались. Такие образцы из дальнейшего анализа исключали.

Таблица 1. Результаты соответствия МПС-аллелей 22 аутосомных STR-локусов, перекрывающихся с PowerPlex® Fusion 6C System

Локус	Количество секвенированных (МПС) аллелей	Количество конкордантных аллелей	Конкордантность, %
<i>Amelogenin</i>	324	324	100
<i>D3S1358</i>	324	324	100
<i>vWA</i>	324	324	100
<i>D16S539</i>	324	324	100
<i>CSF1PO</i>	324	324	100
<i>TPOX</i>	324	324	100
<i>D8S1179</i>	324	324	100
<i>D21S11</i>	324	324	100
<i>D18S51</i>	324	324	100
<i>D2S441</i>	324	324	100
<i>D19S433</i>	324	324	100
<i>TH01</i>	324	324	100
<i>FGA</i>	324	324	100
<i>D22S1045</i>	320	320	100
<i>D5S818</i>	324	322	99.38
<i>D13S317</i>	324	324	100
<i>D7S820</i>	324	323	99.69
<i>D10S1248</i>	324	324	100
<i>DIS1656</i>	317	317	100
<i>D12S391</i>	324	324	100
<i>D2S1338</i>	324	324	100
<i>PentaD</i>	324	324	100
<i>PentaE</i>	322	322	100

Исследование на конкордантность генотипирования

Соответствие между генотипами, установленными капиллярным электрофорезом (КЭ-генотипами) и генотипами, полученными методом МПС (МПС-генотипами), было исследовано для 165 образцов популяционной выборки. Для трех образцов ДНК МПС-генотипы в локусах *DIS1656*, *PentaE* и *PentaD* не регистрировались, эти образцы были исключены из сравнительного анализа. Результаты исследования конкордантности МПС- и КЭ-генотипирования для 162 образцов ДНК представлены в табл. 1.

Дискордантность результатов генотипирования наблюдали в 16 из 3564 сравнений (0.45%), при этом все случаи несоответствия были установлены для пяти локусов: *DIS1656*, *D22S1045*, *D5S818*, *D7S820*, *PentaE*. Для локусов *DIS1656*, *D22S1045*, *PentaE* все случаи несоответствия генотипов были связаны с “выпадением” более длинного аллеля гетерозиготного МПС-профиля, в локусе *D5S818* во всех случаях выпадал более короткий аллель. При анализе файлов отчетов ForenSeq

UAS “выпадающие” аллели обнаруживали, однако уровень прочтений для них был ниже установленного ИП (<30 ридов) (рис. 1,а). Больше всего случаев “выпадения” аллелей зарегистрировано для локуса *DIS1656* (7 образцов), при этом в пяти из семи случаев это были аллели вида X.3, включающие тринуклеотид [TCA]. Согласно [17] аллели X.3 в локусе *DIS1656* значительно хуже секвенируются по сравнению с аллелями без тринуклеотида, что приводит к высокой несбалансированности гетерозигот вида X/X.3.

В локусе *D22S1045* наблюдали дискордантность КЭ- и МПС-генотипов для четырех образцов: три образца имели гомозиготные МПС-генотипы “15/15” и один – “11/11”, тогда как по данным генотипирования тест-системой PowerPlex® Fusion 6C для этих образцов были установлены гетерозиготные профили “15/19” и “11/15” соответственно (рис. 1,б). Несбалансированная амплификация локуса *D22S1045* является известной особенностью тест-системы ForenSeq™, и производитель прямо указывает на необходи-

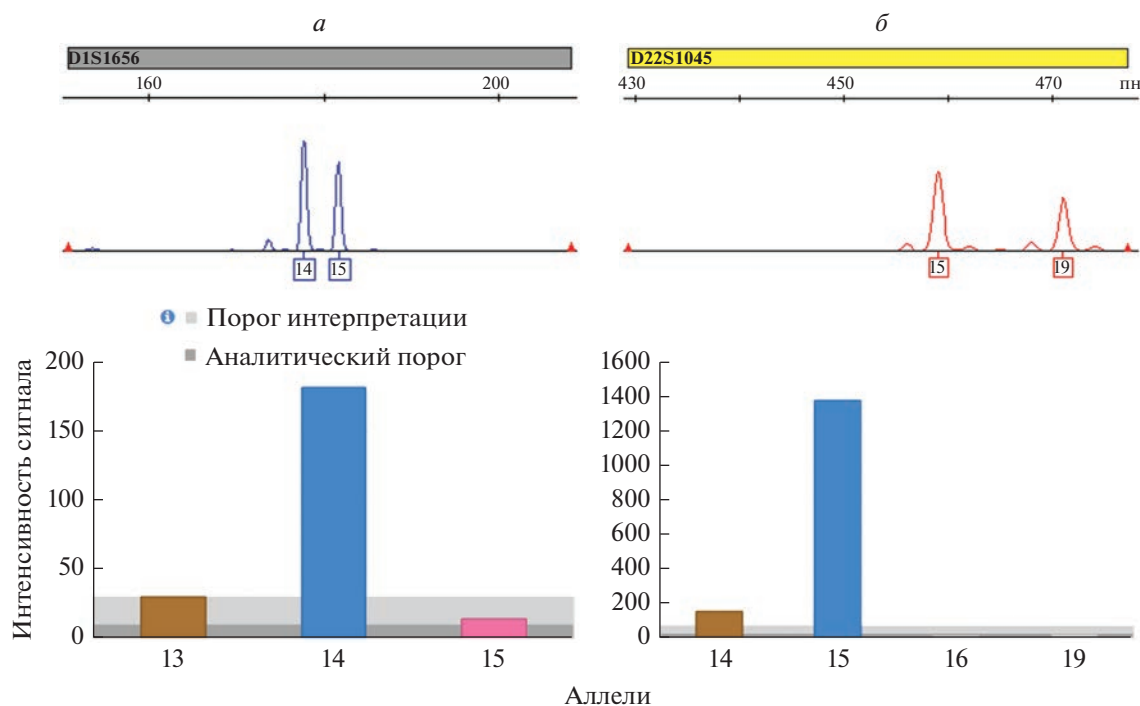


Рис. 1. Результаты генотипирования локусов *DIS1656* и *D22S1045* методами КЭ (вверху) и МПС (внизу) при исследовании конкордантности. *a* – локус *DIS1656*, КЭ-генотип “14/15”, МПС-генотип “14/14”; *б* – локус *D22S1045*, КЭ-генотип “14/19”, МПС-генотип “15/15”.

мость с осторожностью интерпретировать гомозиготные профили по этому маркеру [10].

В локусе *D7S820* для одного из образцов был зарегистрирован КЭ-генотип “10/10.3” и МПС-генотип “10/11”. В локусе *D5S818* несоответствие генотипов обнаружено в двух образцах (КЭ-аллели: “9/11”, МПС: “11” и КЭ: “12/13”, МПС: “13”) и в обоих случаях для незарегистрированных программным обеспечением аллелей (9 и 12) обнаруживались прочтения ниже ИП (<30 ридов).

Наблюдаемые случаи несовпадения генотипов для локусов *D5S818* и *D7S820* были аналогичны описанным ранее наблюдениям [13, 18, 19]. Например, L. Devesse et al. [13] для локуса *D7S820* наблюдали КЭ-аллель 9.2, состоящий из 10 тетра-нуклеотидных повторов, и аллель 7, который был генотипирован как 6.3 на основе методов КЭ. Авторами установлено, что в обоих случаях расхождения связаны с делециями во фланкирующей области – двух оснований на фланке аллеля 10 и одного основания на фланке аллеля 7 (редкая делеция rs540346880, [T/–], обнаруженная в европеоидной популяции с частотой менее 0.01%) (рис. 2,а). В этом же исследовании описан случай “выпадения” короткого аллеля в локусе *D5S818* (по данным КЭ: “9/11”; по данным МПС: “11”). Путем секвенирования более длинного ампликона L. Devesse et al. [13] удалось обнаружить редкую, не описанную ранее однонуклеотидную за-

мену G>C, расположенную в сайте связывания обратного праймера, которая снижала успешность амплификации короткого аллеля 9 (рис. 2,б).

Таким образом, нами показано, что между генетическими профилями 162 образцов, установленными параллельно двумя методами, отсутствуют серьезные несоответствия – конкордантность данных составляет 99.96% (табл. 1). Во всех случаях для незарегистрированных аллелей МПС-профилей имелись прочитанные последовательности, размер которых соответствовал данным КЭ-генотипирования, за исключением одного описанного выше случая для локуса *D7S820*. Вместе с тем результаты исследования позволяют акцентировать внимание на особенностях интерпретации данных, получаемых методом МПС с использованием коммерческого набора ForenSeq DNA Signature Prep Kit [10]. В частности, исследователи должны с особой осторожностью интерпретировать результаты секвенирования не только для локуса *D22S1045*, на что указывает производитель набора, но и для локусов *D5S818*, *DIS1656* и *D7S820*. В соответствии с рекомендациями Национального исследовательского совета США (NRC II Recommendations for Estimating Random-Match Probabilities) [20] чтобы компенсировать эффекты возможных нулевых аллелей в локусах *D5S818*, *DIS1656*, *D22S1045* при оценке вероятности случайного совпадения генотипов

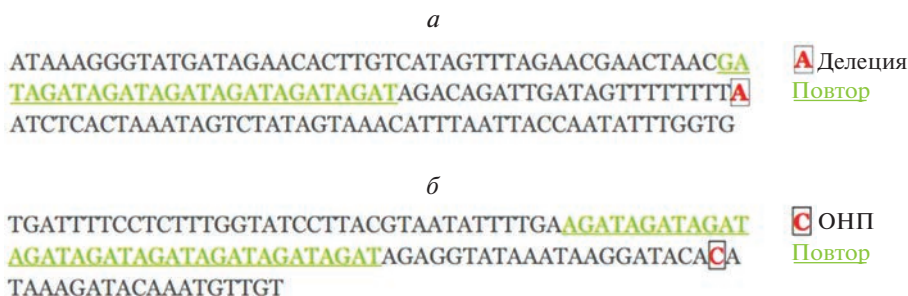


Рис. 2. Мутации фланкирующих последовательностей, приводящие к возникновению “нулевых аллелей” при МПС-генотипировании локусов *D7S820* и *D5D818*. Адаптировано из [13]. *a* – локус *D7S820* (rs540346880); *б* – локус *D5S818*.

может применяться высококонсервативный подход, при котором частоту (f) для гетерозигот рассчитывают как $f = 2pq$, а для гомозигот – как $f = p$ вместо $f = p^2$. Более точная, но также консервативная процедура расчета заключается в использовании уравнения $f = p^2 + p(1 - p) \times \theta$ со значением $\theta = 0.003$ [20].

Генетическое разнообразие и криминалистическая информативность секвенированных аутосомных STR-локусов

Из 39534 аллелей, секвенированных у 733 индивидов в 27 локусах, был идентифицирован 301 вариант различной длины и 517 изометрических вариантов (аллелей одинаковой длины, но разной последовательности).

Число аллелей увеличилось более чем в 2 раза ($\times 2$) у восьми локусов: *D12S391* ($\times 4.24$), *D21S11* ($\times 3.43$), *D2S1338* ($\times 3.07$), *vWA* ($\times 2.63$), *D3S1358* ($\times 2.44$), *D8S1179* ($\times 2.0$), *D13S317* ($\times 2.11$), *D9S1122* ($\times 2.0$). Всего из 27 аутосомных STR-локусов изоаллели наблюдались в 19 (70.37%). Для восьми локусов не зафиксировано дополнительных МПС-аллелей (*PentaE*, *PentaD*, *D20S482*, *CSF1PO*, *TPOX*, *D10S1248*, *TH01*, *D16S539*) (табл. 2). Полученные в нашем исследовании результаты хорошо согласуются с данными других авторов, полученными при исследовании европейских и азиатских популяций [13, 17–19, 21–24].

Для шести локусов (*D19S433*, *D7S820*, *D18S51*, *D5S820*, *D13S317* и *vWA*) дополнительно были проанализированы предоставляемые ForenSeq UAS данные о последовательности небольших фланкирующих участков, непосредственно прилегающих к тандемной области. Для двух локусов *D7S820* и *D18S51* не показано увеличения количества аллелей за счет фланков – варибельность отмечалась исключительно в области повтора. С учетом информации о фланкирующих последовательностях в локусе *D19S433* был идентифицирован редкий изоаллель 13.2 – *D19S433[CE13.2]-Chr19 GRCh37p.13 30417142-30417205 [CCTT]12 ccta CCTT cttt CCTT rs745607776-del*; в локусе *D13S317* – во-

семь изоаллелей с описанными частыми мутациями А>Т в первой и пятой позициях на 3'- фланкирующей области повтора (rs9546005, rs202043589); в локусе *vWA* – четыре изоаллеля: 14 и 15 (rs75219269 А>G), 16, 17 (rs199970098 G>А) с полиморфизмом во фланкирующей области. Тогда как для локуса *D5S818* вся варибельность обеспечивалась известной мутацией во фланкирующем участке (rs73801920 С>А) [24] и только в одном случае был обнаружен описанный ранее вариант аллеля 12, несущий мутацию (rs1171357623, Т>G) в четвертом тандеме [24] (табл. 3).

Из 517 различных аллелей, идентифицированных в белорусской популяции, 13 ранее не были описаны (табл. 4). Среди впервые обнаруженных аллелей семь были обусловлены новой комбинацией повторяющихся блоков в тандемной области, пять имели однонуклеотидные замены внутри тандемного повтора и один аллель – 15 в локусе *D5S818* имел на фланке мутацию rs73801920, описанную ранее, но в комбинации с этим вариантом аллеля не встречавшуюся в других исследованных популяциях, согласно сведениям STRBase 2.0 [7]. Все ранее не описанные аллели зарегистрированы в базе данных Nucleotide NCBI под номерами ON248964, ON359896–ON359907.

При оценке уровня генетического разнообразия белорусской популяции с учетом данных МПС-генотипирования показано увеличение значений наблюдаемой гетерозиготности (H_o) для всех локусов с выявленными изоаллельными вариантами, кроме локусов *FGA*, *D17S1301*, *D18S51*, *D22S1045*. В наибольшей степени H_o увеличивалась для локуса *D4S2408* – в 1.32 раза. Для этого локуса выявлен полиморфизм нуклеотидной последовательности области тандемного повтора только для аллеля 9, имеющего максимальную частоту встречаемости в популяции (59.5%), при этом частота встречаемости двух выявленных изоморфных вариантов оказалась практически одинаковой – 0.2196 и 0.1869. Таким образом, из 261 образца, имевшего по данным КЭ-анализа гомозиготный генотип “9/9”, 137 образцов, т.е. более

Таблица 2. Число STR-аллелей, идентифицированных методом капиллярного электрофора (КЭ-аллели) и массовым параллельным секвенированием (МПС-аллели)

Локус	КЭ-аллели	МПС-аллели	Прирост, <i>n</i> (%)
<i>D12S391</i>	17	72	+ 55 (324)
<i>D21S11</i>	15	48	+ 33 (220)
<i>D2S1338</i>	14	43	+ 29 (207)
<i>vWA</i>	8	21	+ 13 (163)
<i>D3S1358*</i>	9	22	+ 13 (144)
<i>D13S317</i>	9	19	+ 10 (111)
<i>D8S1179</i>	11	22	+ 11 (100)
<i>D9S1122</i>	6	12	+ 6 (100)
<i>D5S818*</i>	9	17	+ 8 (89)
<i>D2S441</i>	9	13	+ 4 (44)
<i>D1S1656</i>	19	28	+ 9 (47)
<i>D6S1043</i>	16	23	+ 7 (44)
<i>D19S433</i>	16	22	+ 6 (38)
<i>D7S820</i>	8	10	+ 2 (25)
<i>FGA</i>	19	23	+ 4 (21)
<i>D18S51</i>	15	18	+ 3 (20)
<i>D4S2408</i>	7	8	+ 1 (14)
<i>D17S1301</i>	8	9	+ 1 (13)
<i>D22S1045</i>	8	9	+ 1 (13)
<i>PentaE</i>	19	19	0
<i>PentaD</i>	14	14	0
<i>D20S482</i>	9	9	0
<i>CSF1PO</i>	8	8	0
<i>TPOX</i>	7	7	0
<i>D10S1248</i>	7	7	0
<i>TH01</i>	7	7	0
<i>D16S539</i>	7	7	0

* Вариабельность преимущественно обеспечена мутациями на фланках.

половины, имеют гетерозиготный генотип. Похожая картина наблюдалась для локусов *D9S1122* и *D13S317*, для которых увеличение гетерозиготности по сравнению с данными КЭ-генотипирования также было значительным (табл. 5).

Вероятность случайного совпадения генотипов двух неродственных индивидов (*MP*) снижалась для всех локусов, для которых были зарегистрированы изоаллельные варианты (кроме локуса *FGA*), что свидетельствует об увеличении идентификационной значимости результатов генотипирования, получаемых методом МПС. В наибольшей степени *MP* снижалась для локусов *D9S1122*, *D8S1179*, *D13S317* и *D21S11* – в 2.7 раза; в наименьшей степени – для локусов *D22S1045* и *D17S1301* (в 1.1 раз). Рассчитанная по данным МПС средняя вероятность того, что два случайно выбранных из популяции индивида будут иметь

один и тот же генотип по совокупности всех 27 STR-локусов, составила 2.89×10^{-35} (это означает, что случайное совпадение генотипа возможно с одним из 3.5×10^{34} человек). Этот результат на четыре порядка превышает значение, получаемое с использованием данных фрагментного анализа (1.43×10^{-31} , один из 7×10^{30} человек).

Информационная значимость данных МПС-генотипирования при установлении родства также значительно увеличивается по сравнению с данными КЭ, о чем свидетельствует рост значимости вероятности исключения индивида по заданному генотипу (*PE*) и индексов родства (*TPI*) для локусов тест-системы. Так, *PE* увеличивалась в наибольшей степени для локуса *D4S2408* – в 2 раза, для локусов *D9S1122*, *D13S317*, *D5S818* и *D8S1179* – в 1.3–1.5 раза. Максимальный рост значений *TPI* зафиксирован для локусов *D12S391*

Таблица 3. Аллели с мутациями во фланкирующих последовательностях (*N* – количество аллелей в популяции)

Локус	Аллель		<i>N</i>	Локус	Аллель		<i>N</i>
<i>D5S818</i>	9а	[ATCT]9_rs73801920-A	1	<i>D13S317</i>	9б	[TATC]9_rs9546005-T	1
	10б	[ATCT]10_rs73801920-A	28		9в	[TATC]10_rs9546005-T	3
	11б	[ATCT]11_rs73801920-A	55		10в	[TATC]10_rs9546005-T, rs202043589-T	1
	12б	[ATCT]12_rs73801920-A	137		11а	[TATC]11_rs9546005-T	272
	12в	[ATCT]3 AGCT [ATCT]8_rs73801920-A	1		12б	[TATC]12_rs9546005-T	163
	13б	[ATCT]13_rs73801920-A	57		13б	[TATC]13_rs9546005-T	46
	14а	[ATCT]14_rs73801920-A	8		14б	[TATC]14_rs9546005-T	8
	15б	15_[ATCT]15_rs73801920-A	1		15а	[TATC]15_rs9546005-T	1
Локус	Аллель						<i>N</i>
<i>vWA</i>	14а	[TAGA]3 TGGA [TAGA]3 [CAGA]4 TAGA CAGA TAGA_rs75219269-G					127
	15б	TGGA [TAGA]3 TGGA [TAGA]3 [CAGA]4 TAGA CAGA TAGA_rs75219269-G					2
	16в	[TAGA]11 [CAGA]4 TAGA_rs199970098-A					1
	17д	[TAGA]12 [CAGA]4 TAGA_rs199970098-A					2
<i>D19S433</i>	13.2б	[CCTT]12 ccta CCTT cttt CCTT_rs745607776-delCT					1

Таблица 4. Новые аллели, обнаруженные в белорусской популяции

Локус	Аллель		Число	Покрытие
<i>D2S1338</i>	21г	D2S1338[CE21]-Chr2-GRCh37.p13 218879582-218879673 GGAG [GGAA]13 [GGCA]7	1	1509
<i>FGA</i>	23.2б	FGA[CE23.2]-Chr4-GRCh37.p13 155508888-155508975 [GGAA]2 GGAG AAAG AG [AAAG]14 AGAA AAAA [GAAA]3	1	1333
<i>D5S818</i>	15б	D5S818[CE15]-Chr5-GRCh37.p13 123111250-123111293 [ATCT]15 rs73801920-A	1	209
<i>D7S820</i>	11б	D7S820[CE11]-Chr7-GRCh37.p13 83789542-83789593 [TATC]2 TACC [TATC]8	1	140
<i>D12S391</i>	18.3б	D12S391[CE18.3]-Chr12-GRCh37.p13 12449954-12450029 AGAT GAT [AGAT]10 [AGAC]6 AGAT	3	>280
	23ж	D12S391[CE23]-Chr12-GRCh37.p13 12449954-12450029 [AGAT]16 [AGAC]6 AGAT	2	>200
	26б	D12S391[CE26]-Chr12-GRCh37.p13 12449954-12450029 [AGAT]15 [AGAC]11	1	331
<i>D18S51</i>	16б	D18S51[CE16]-Chr18-GRCh37.p13 60948900-60948971 AGAA AGCA [AGAA]14	1	403
<i>D19S433</i>	11б	D19S433[CE11]-Chr19-GRCh37.p13 30417142-30417205 [CCTT]9 cctt CCTT cttt CCTT	1	1375
	13б	D19S433[CE13]-Chr19-GRCh37.p13 30417142-30417205 ACTT [CCTT]10 ccta CCTT cttt CCTT	1	845
<i>D21S11</i>	29б	D21S11[CE29]-Chr21-GRCh37.p13 20554291-20554417 [TCTA]2 TATA [TCTA]3 [TCTG]5 [TCTA]3 ta [TCTA]3 tca [TCTA]2 tccata [TCTA]10	1	1056
	30в	D21S11[CE30]-Chr21-GRCh37.p13 20554291-20554417 [TCTA]6 [TCTG]5 [TCTA]3 ta [TCTA]2 tca [TCTA]2 tccata [TCTA]12	2	>500
	30ж	D21S11[CE30]-Chr21-GRCh37.p13 20554291-20554417 [TCTA]4 [TCTG]6 [TCTA]4 ta [TCTA]3 tca [TCTA]2 tccata [TCTA]11	1	1071

Таблица 5. Изменение уровня генетического полиморфизма и параметров информационной значимости ауто-сомных STR-локусов, исследованных методом МПС

Локус	H_O		MP		PE		TPI	
	КЭ	МПС	КЭ	МПС	КЭ	МПС	КЭ	МПС
<i>D1S1656</i>	0.889	0.896	0.020	0.016	0.773	0.787	4.51	4.81
<i>D2S441</i>	0.738	0.769	0.100	0.082	0.490	0.544	1.91	2.17
<i>D2S1338</i>	0.855	0.891	0.030	0.018	0.706	0.777	3.46	4.58
<i>D3S1358</i>	0.789	0.857	0.074	0.034	0.578	0.708	2.36	3.49
<i>D4S2408</i>	0.585	0.772	0.230	0.107	0.274	0.548	1.21	2.19
<i>FGA</i>	0.879	0.879	0.037	0.037	0.752	0.752	4.12	4.12
<i>D5S818</i>	0.724	0.808	0.120	0.060	0.467	0.613	1.81	2.60
<i>D6S1043</i>	0.849	0.850	0.049	0.048	0.692	0.695	3.30	3.33
<i>D7S820</i>	0.786	0.793	0.066	0.065	0.573	0.585	2.33	2.41
<i>D8S1179</i>	0.784	0.874	0.080	0.030	0.571	0.744	2.32	3.98
<i>D9S1122</i>	0.707	0.834	0.156	0.058	0.439	0.663	1.70	3.00
<i>vWA</i>	0.802	0.840	0.064	0.042	0.603	0.676	2.53	3.13
<i>D12S391</i>	0.898	0.951	0.024	0.010	0.791	0.900	4.89	10.18
<i>D13S317</i>	0.774	0.883	0.080	0.030	0.551	0.760	2.21	4.26
<i>D17S1301</i>	0.701	0.701	0.162	0.161	0.430	0.430	1.67	1.67
<i>D18S51</i>	0.847	0.847	0.032	0.031	0.689	0.689	3.27	3.27
<i>D19S433</i>	0.779	0.782	0.079	0.077	0.561	0.566	2.26	2.29
<i>D21S11</i>	0.872	0.919	0.040	0.015	0.738	0.835	3.89	6.20
<i>D22S1045</i>	0.724	0.724	0.110	0.109	0.467	0.467	1.81	1.81
<i>TPOX</i>	0.587	0.587	0.228	0.228	0.275	0.275	1.21	1.21
<i>CSFIPO</i>	0.741	0.741	0.115	0.115	0.494	0.494	1.93	1.93
<i>D10S1248</i>	0.768	0.768	0.101	0.101	0.541	0.541	2.16	2.16
<i>D20S482</i>	0.738	0.738	0.134	0.134	0.490	0.490	1.91	1.91
<i>TH01</i>	0.733	0.733	0.091	0.091	0.480	0.480	1.87	1.87
<i>PentaE</i>	0.898	0.898	0.020	0.020	0.791	0.791	4.91	4.91
<i>PentaD</i>	0.809	0.809	0.054	0.054	0.616	0.616	2.62	2.62
<i>D16S539</i>	0.745	0.745	0.101	0.101	0.501	0.501	1.96	1.96
Комбинированное значение			1.43×10^{-31}	2.89×10^{-35}			2.08×10^{10}	3.25×10^{12}

Примечание. Полужирным шрифтом выделены локусы, для которых не зафиксировано дополнительных МПС-аллелей.

(с 4.89 до 10.18) и *D21S11* (с 3.89 до 6.2), т.е. для локусов с наибольшим количеством зарегистрированных по данным МПС аллельных вариантов. Комбинированный индекс родства, получаемый с учетом наличия МПС-аллелей, достигал значения 3.25×10^{12} , что на два порядка выше значения, получаемого на основе данных фрагментного анализа — 2.08×10^{10} .

Анализ генетической подразделенности резидентного населения Республики Беларусь, проведенный методом молекулярной дисперсии (AMOVA) с учетом различий в частотах МПС-аллелей, не выявил статистически значимых генетических различий между исследованными региональными популяциями. Все население рес-

публики характеризуется общностью генного пула по изученным 27 аутосомным STR-маркерам. Анализ матриц парных расстояний также не позволил выявить статистически достоверные различия между парами популяционных выборок, сформированных по региональному признаку. Применительно к экспертному ДНК-анализу полученные данные указывают на возможность использования частотных характеристик 27 аутосомных локусов, установленных методом МПС, для всего населения Беларуси и формирования единой республиканской референсной базы данных.

Частоты аллелей аутосомных STR-локусов для населения Республики Беларусь по данным массового параллельного секвенирования приведены в

дополнительных материалах к настоящей статье и могут быть использованы для вероятностных расчетов при идентификации личности и установлении родства в криминалистике, а также в популяционно-генетических сравнительных исследованиях.

В рамках настоящего исследования с использованием технологии МПС и набора ForenSeq DNA Signature Prep Kit впервые у населения Республики Беларусь изучена внутренняя структура нуклеотидной последовательности 27 аутосомных STR-локусов. Проведена оценка соответствия результатов генотипирования, получаемых методом капиллярного электрофореза и массовым параллельным секвенированием. Показано, что генотипы PowerPlex® Fusion 6C и ForenSeq DNA Signature Prep Kit конкордантны для всех перекрывающихся локусов, за исключением *D7S820* — соответствие составило 99.96%. Вместе с тем следует отметить, что для отдельных образцов в некоторых локусах обнаружены ложные гомозиготные генотипы, один из аллелей в которых присутствовал, но не детектировался программным обеспечением ForenSeq UAS. В связи с этим при проведении исследования с использованием набора ForenSeq DNA Signature Prep Kit рекомендовано: 1) особое внимание обращать на результаты генотипирования локусов *D22S1045* и *D7S820*; также требуется осторожность в отношении локусов *D5S818*, *D1S1656*, *PentaE* и *PentaD*, особенно при исследовании малых количеств ДНК или ДНК плохого качества; 2) на основе анализа эмпирических данных определить минимальные пороговые уровни прочтения маркеров, преимущественно маркеров с высоким гетерозиготным дисбалансом; 3) установить оптимальные требования к качеству исследуемых образцов.

С учетом полиморфизма тандемных и фланкирующих участков исследованных локусов в белорусской популяции суммарно идентифицировано 216 дополнительных аллельных вариантов, 13 из которых обнаружены впервые. Для всех локусов рассчитаны частоты встречаемости аллелей и криминалистические параметры идентификационной информативности. Для имплементации новых сведений в экспертную практику сформирована референтная база данных частот аутосомных STR-локусов для населения Республики Беларусь по данным массового параллельного секвенирования.

Полученные в настоящей работе сведения предоставляют для экспертного сообщества Республики Беларусь возможность проведения вероятностной оценки результатов идентификации личности и установления родства, основанную на исследовании МПС-профилей ДНК, что обеспечит эффективное и корректное использование методов массового параллельного секвенирования в экспертной практике.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национально-го комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Животовский Л.А.* Микросателлитная изменчивость в популяциях человека и методы ее изучения // Вестник ВОГиС. 2006. Т. 10. № 1. С. 74–96.
2. *van der Gaag K.J., de Leeuw R.H., Hoogenboom J. et al.* Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq™ system // Forensic Sci. Intern.: Genetics. 2016. V. 24. P. 86–96. <https://doi.org/10.1016/j.fsigen.2016.05.016>
3. Thermo Fisher Scientific [Электронный ресурс]. URL: <https://www.thermofisher.com/by/en/home.html> (дата обращения 03.04.2022).
4. Illumina. Sequencing and array-based solutions for genetic research [Электронный ресурс]. URL: <https://www.illumina.com/> (дата обращения 03.04.2022).
5. *Цыбовский И.С., Веремейчик В.М., Котова С.А. и др.* Создание судебной референтной базы данных по 18 аутосомным STR для ДНК-идентификации в Республике Беларусь // Генетика. 2017. Т. 53. № 2. С. 249–258.
6. *Харьков В.Н., Котова С.А., Колесников Н.А. и др.* Генетическое разнообразие 21 аутосомного STR-маркера системы CODIS в популяциях Восточной Европы // Генетика. 2021. Т. 57. № 12. С. 1396–1402.
7. STRbase 2.0 [Электронный ресурс]. URL: <https://strbase-b.nist.gov> (дата обращения 03.04.2022).
8. *Chacon-Cortes D., Haupt L.M., Lea R.A. et al.* Comparison of genomic DNA extraction techniques from whole blood samples: A time, cost and quality evaluation study // Mol. Biol. Rep. 2012. V. 39. № 5. P. 5961–5966. <https://doi.org/10.1007/s11033-011-1408-8>
9. *Маниатис Т., Фрич Э., Сэмбрук Дж.* Методы генетической инженерии. Молекулярное клонирование. М.: Мир, 1984. 478 с.
10. Verogen, ForenSeq™ DNA Signature Prep Reference Guide // Verogen proprietary. Document # VD2018005 Rev. A, 2018. 42 p.
11. GRCh37.p13: Genome Reference Consortium Human Build 37 patch release 13 [Электронный ресурс]. URL: https://www.ncbi.nlm.nih.gov/assembly/GCF_0000014-05.25 (дата обращения 03.02.2021).
12. *Parson W., Ballard D., Budowle B. et al.* Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements // Forensic Sci. Intern.: Genetics. 2016.

- V. 22. P. 54–63.
<https://doi.org/10.1016/j.fsigen.2016.01.009>
13. *Devesse L., Ballard D., Davenport L. et al.* Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups // *Forensic Sci. Intern.: Genetics*. 2018. V. 34. P. 57–61.
<https://doi.org/10.1016/j.fsigen.2017.10.012>
 14. GeneAIEx 6.5: Genetic Analysis in Excel [Электронный ресурс]. URL: <https://biology-assets.anu.edu.au/GeneAIEx/Welcome.html> (дата обращения 15.11.2021).
 15. *Ristow P.G., D'Amato M.E.* Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics // *Forensic Sci. Intern.: Genetics Suppl. Series 6*. 2017. V. 6. P. e52–e54.
<https://doi.org/10.1016/j.fsigs.2017.09.006>
 16. Arlequin: An Integrated Software for Population Genetics Data Analysis [Электронный ресурс]. URL: <http://cmpg.unibe.ch/software/arlequin3> (дата обращения 02.10.2018).
 17. *Hussing C., Huber C., Bytyci R. et al.* Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system – an evaluation of the assay and software // *Forensic Sci. Research*. 2018. V. 3. №. 2. P. 111–123.
<https://doi.org/10.1080/20961790.2018.1446672>
 18. *Dai W., Pan Y., Sun X. et al.* High polymorphism detected by massively parallel sequencing of autosomal STRs using old blood samples from a Chinese Han population // *Sci. Rep.* 2019. V. 9. P. 1–7.
<https://doi.org/10.1038/s41598-019-55282-9>
 19. *Hözl-Müller P., Bodner M., Berger B. et al.* Exploring STR sequencing for forensic DNA intelligence databasing using the Austrian National DNA Database as an example // *Int. J. Legal Med.* 2021. V. 135. P. 2235–2246.
<https://doi.org/10.1007/s00414-021-02685-x>
 20. *The Evaluation of Forensic DNA Evidence. Committee on DNA Forensic Science: an update.* Washington (D.C.): Natl Acad. Press, 1996. 272 p.
<https://doi.org/10.17226/5141>
 21. *Gettings K.B., Aponte R.A., Vallone P.M. et al.* STR allele sequence variation: Current knowledge and future issues // *Forensic Sci. Intern.: Genetics*. 2015. V. 18. P. 118–130.
<https://doi.org/10.1016/j.fsigen.2015.06.005>
 22. *Gettings K.B., Borsuka L.A., Steffen C.R. et al.* Sequence-based U.S. population data for 27 autosomal STR loci // *Forensic Sci. Intern.: Genetics*. 2018. V. 37. P. 106–115.
<https://doi.org/10.1016/j.fsigen.2018.07.013>
 23. *Simayijiang H., Morling N., Børsting C.* Sequencing of human identification markers in an Uyghur population using the MiSeq FGx™ Forensic Genomics System // *Forensic Sci. Research*. 2020. P. 1–9.
<https://doi.org/10.1080/20961790.2020.1779967>
 24. *Gettings K.B., Borsuk L.A., Ballard D. et al.* STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci // *Forensic Sci. Intern.: Genetics*. 2017. V. 31. P. 111–117.
<https://doi.org/10.1016/j.fsigen.2017.08.017>

Variability of 27 Autosomal STR Loci for the Population of the Republic of Belarus Based on the Mass Parallel Sequencing Data

S. A. Kotova^{a, *}, N. S. Parfionava^a, T. V. Zabauskaya^a,
 V. I. Rybakova^a, A. A. Spivak^a, S. A. Paliavoi^b, and A. V. Lugovnev^c

^aScientific and Practical Center of the State Forensic Examination Committee of the Republic of Belarus, Minsk, 220114 Republic of Belarus

^bMasaryk University, Brno, 60177 Czech Republic

^cState Forensic Examination Committee of the Republic of Belarus, Minsk, 220073 Republic of Belarus

*e-mail: svetlkotova@mail.ru

Variability of 27 autosomal STR loci of the ForenSeq DNA Signature Prep Kit (Illumina) commercial panel was studied using the technology of mass parallel sequencing (MPS) in 733 unrelated individuals representing the population of the Republic of Belarus as well as a population base of MPS allele frequencies for expert probabilistic calculations in human identification and paternity establishment was evaluated. The agreement between genotypes obtained by MPS and capillary electrophoresis (CE) was 99.96%. The number of MPS alleles increased more than two times for eight loci (*D12S391*, *D21S11*, *D2S1338*, *vWA*, *D3S1358*, *D8S1179*, *D13S317*, *D9S1122*). Thirteen alleles detected were not included in the STRSeq BioProject catalog of the international online database STRbase 2.0. The random match probability of 27-locus MPS profiles decreased from 1.43×10^{-31} to 2.89×10^{-35} , and the combined paternity index increased from 2.08×10^{10} to 3.25×10^{12} compared to CE data.

Keywords: autosomal STR locus, variability, isometric alleles, massively parallel sequencing.