**Supplemental Material**

**Bioinformatical analysis of the 1000 Genomes (1000G) phase 3 data**

In a first step, 1000 Genomes project [s1] phase 3 data were downloaded from ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ (date of accession: 20 March 2015; EBI later replaced/updated/renamed these files on 18 August 2015). Structural variant (SV) calls and affected sample IDs were extracted from the file ALL.chr4.phase3\_shapeit2\_mvncall\_integrated\_v5a.20130502.genotypes.vcf for genomic region chr4:144792019-145061904 in hg19-coordinates. This genomic region contains the genes *GYPE*, *GYPB* and *GYPA*. SV frequencies in populations and subpopulations were determined after assignment of sample IDs to populations and subpopulations with the help of file integrated\_call\_samples\_v3.20130502.ALL.panel and a self written python script. In a second step, each SV call in each affected sample was subjected to manual validation using the Integrative genome viewer (IGV) [s2, s3]. For this purpose, the BAM file of sequence alignments was downloaded for each sample ID from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/<sample-id>/alignment/. Potentially true large deletion candidates in the region of interest were confirmed if (a) paired-end reads were found with an extremely large separation distance after alignment, especially if a soft-clipped part of one end could be mapped in the region of the other end, or (b) distinct coverage drops were found, or (c) complete coverage gaps were found, or (d) a combination of several of these criteria was true. New large deletion candidates were visually identified in the BAM files using the same criteria (a)-(d). A UCSC BLAT [s4] search for matching loci on the human reference sequences hg19 and hg38 was used to assess the specificity of aligned paired-ends that spanned a potentially true deletion.

**Other analyses**

Homology analysis of the *GYP* locus started with unit *GYPA* at genomic coordinate 144147912 of NC\_000004.12 (GRCh38.p12, Primary Assembly), with a sequence reading GAAAAGCTACTCCTGCTTGC. Alignments were done using GeneRunner, version 5.1.0.6.Beta, Copyright 1992-2016 Frank Buquicchio [s5-s7]. Alignment details are given in the supplementary table s6.

Allele frequency calculations were done as described previously [s8].

**Molecular Typing Hierarchy**

Due to acute DNA-depletion of some samples, genotyping of all samples followed a strictly defined typing hierarchy, allowing the early termination of unnecessary further genetic analysis (“waste of DNA”), whenever final genotypes, as defined by an unambiguous identification of two different parental alleles (haplotypes) occurred:

Step 1. Typing of SNPs *GYPA* c.59C>T (M/N, rs7682260) and *GYPB* c.143C>T (S/s, rs7683365). This typing was mandatory for all samples. All DNAs of panel A originating from Milwaukee, Vienna, Hagen, Lund and the Coriell Cell Repository (n = 34) were typed with the commercially available PCR-SSP based typing kit MNS and according to the manufacturers’ instructions (Inno-Train GmbH, Kronberg i.T., Germany). Previously, all other DNAs of panel A (n = 82) had been genotyped for both SNPs using MALDI-TOF MS [s8]. All 157 DNAs of panel B were freshly analysed using MALDI-TOF [s9].

Step 2. Specific detection of the two distinct *GYPB* deletions with either a length of 110 kb, or 103 kb using Gap-PCRs (reactions 110-1.3, and 103-1.3, suppl. table s1C). To discriminate homo- versus heterozygosity for samples with *GYPB* deletions, all samples with positivity for either the 110 kb or the 103 kb deletion, were additionally tested for the presence of the undeleted wildtype haplotypes, e.g. lacking a *GYPB* deletion (reactions 110-1.1 and 110-1.2, plus 103-1.1 and 103-1.2, suppl. table s1C).

Step 3. Supplementation of all results for *GYPB* SNP *GYPA* c.59T>G (He) already available from MALDI-TOF MS was done using PCR-SSP based genotyping (reactions “He” and “non He”, suppl. table s1C). All *GYPA* c.59G positives were further tested for the presence and zygosity of U+w (“Uvar”) alleles defined by *GYPB* SNPs c.230C>T and c.270+5g>t, if results were unavailable from MALDI-TOF data (reactions “U+w, NY, He(NY)” and “non U+w, NY, He(NY)” plus “U+w, P2, He(P2)” and “non U+w, P2, He(P2)”, suppl. table s1C).

Step 4. Independently, testing for the two 1000G samples of the Coriell Human Genetic Cell Repository, carrying the expected ~32 kb (repository number: HG01880\*C1) and ~19 kb (repository number: HG01880\*C1) *GYPB* deletions was done by PCR-SSP based analysis as described above, or with reactions specifically developed for the 19 kb *GYPB* deletion (reactions 019-1.1, 019-1.2, 019-1.3).

According to the described hierarchy, some few samples were not tested for *GYPB* c.230C>T and c.270+5g>t, defining status for U+var alleles lacking the “He”-specific *GYPB* SNP c.59T>G. As a consequence, rare alleles G*YPB\*03N.01*, or *GYPB\*03N.03*, if they were present within panel A and *GYPB\*03N.01*, if they were present within panel B would have gone undetected and misinterpreted as regular *GYPB\*03* alleles.

Step 5. Testing for linkage. Hybrid sequences analysed for the 110 kb and 103 kb *GYPB* deletions, showed two and three “private” nucleotides respectively (Figure 2A and 2B), not found in neither of the individual wildtype sequences. Testing all CEPH DNAs for four of these specific nucleotides by PCR-SSP was done using primers given (suppl. table s1D),

.

**Supplementary Table s1, A to C. Primers used, reaction conditions, amplicon sizes and sequencing**

**Section A** gives primers used for the amplification of the three putative breakpoint regions for the regular, e.g. wildtype, haplotypes of all three *GYPA*, *GYPB* and *GYPE* units, and primers used for Gap-PCRs [s10]. “Gap-PCRs”, expected to deliver amplification products only in the presence of one of the two presumptive ~100 kb, or ~19 kb *GYPB* deletions were designed for the expected ~100 kb *GYPB-GYPE* spanning deletion by combining *GYPB* unit-specific forward primers together with *GYPE* unit-specific reverse primers. In analogy, *GYPA* unit-specific forward primers were combined with *GYPB* unit-specific reverse primers for the expected ~100 kb *GYPA-GYPB* spanning deletion. For the expected ~19 kb *GYPB* deletion, gene internal forward primers starting in intron 1 were combined with reverse primers of the 3’ end of the gene. All amplifications expected from the paralogous breakpoint candidate regions involving the two distinct ~100 kb *GYPB* deletions were carried out in duplicate (“long” and “short”). **Section B** shows primers used for sequencing of the amplification products resulting from section A. **Section C** displays a set of assays consisting of Gap-PCRs and PCRs using sequence-specific priming technique (PCR-SSP) [s11, s12], specifically designed to classify allelic *GYPB* SNPs and hybrid-haplotypes (deemed ”diagnostic PCR-SSPs”) on the basis of the deletions characterized. **Section D** shows primers specifically designed to detect “private” nucleotides explained further below. **Section E** shows primers used as positive amplification control in all reactions detailed in sections C and D.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **Types of Primers** Reaction Numbers |  |  | **Intentended Usage** |  | **Primer Name** |  | **Primer Sequence 5' -> 3'** | **Concen-tration (nM)** | **Amplifica-tion Control** | **Amplicon  Size (bp)** | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **A. Gap-PCR Amplicons for Sequencing Breakpoints** |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 110-3.1 |  |  | "short" for *GYPA* wildtypes |  | GYPA-110k-448-j5-F |  | AGGGATGTGGCTTCCTGACGG | 300 | no | 1659 | |
|  |  |  |  |  | GYPA-110k+740-j3-R |  | CATCACTACAATGGTGACGTAGACAGTGATTA | 300 |  |  | |
| 110-3.2 |  |  | "short" for *GYPB* wildtypes |  | GYPB-110k-448-j5F |  | CTAGGGATGTGGCTTCCTGACGA | 300 | no | 1293 | |
|  |  |  |  |  | GYPB-110k+379-j3R |  | GAAAGGTTTTCCAGAGCAAGCAAGA | 300 |  |  | |
| 110-3.3 |  |  | "short" for *GYPE* wildtypes |  | GYPE-110k-299-j5-F |  | TGCAACAGCACCGAGTCTGTTACC | 300 | no | 1184 | |
|  |  |  |  |  | GYPE-110k+414-j3-R |  | CATAAGAGGGGTACCTAACTCAGGGTTGA | 300 |  |  | |
| 110-3.4 |  |  | "short" for 110 kb *GYPB* deletion |  | GYPB-110k-272-j5-F |  | TGGGTGAGCAGGGCTGAGATTC | 300 | no | 737 | |
|  |  |  |  |  | GYPE-110k+414-j3-R |  | CATAAGAGGGGTACCTAACTCAGGGTTGA | 300 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 110-4.1 |  |  | "long" for *GYPA* wildtypes |  | GYPA-110k-1797-F |  | GTTTAAGGAGGCTGAAGACAGGGACA | 300 | no | 2871 | |
|  |  |  |  |  | GYPA-110k+1022-R |  | CAAATGGAAGTCCTGCAAGTCATCGT | 300 |  |  | |
| 110-4.2 |  |  | "long" for *GYPB* wildtypes |  | GYPB-110k-1705-F |  | TTATAGGTTAACTCGTCCTTTTACCGCAA | 300 | no | 2827 | |
|  |  |  |  |  | GYPB-110k+1071-R |  | CAGTGGCCAAGCTGACCTCGAC | 300 |  |  | |
| 110-4.3 |  |  | "long" for *GYPE* wildtypes |  | GYPE-110k-1647-F |  | TTTAGATAACCTGATGACAATGTGCGTG | 600 | no | 2439 | |
|  |  |  |  |  | GYPE-110k+737-R |  | TCACTACAATGGTGATGTAGACAGCGG | 600 |  |  | |
| 110-4.4 |  |  | "long" for 110 kb *GYPB* deletion |  | GYPB-110k-1705-F |  | TTATAGGTTAACTCGTCCTTTTACCGCAA | 600 | no | 2498 | |
|  |  |  |  |  | GYPE-110k+737-R |  | TCACTACAATGGTGATGTAGACAGCGG | 600 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 103-5.1 |  |  | "short" for *GYPA* wildtypes |  | GYPA-103k-440-j5F |  | CACAAACTCAGTAATTCTCTTTAGAGTGACGC | 300 | no | 1374 | |
|  |  |  |  |  | GYPA-103k+873-j5R |  | ACATGATGACAGAAGTAAATGGGGAAGAG | 300 |  |  | |
| 103-5.2 |  |  | "short" for *GYPB* wildtypes |  | GYPB-103k-688-j3F |  | AGCTTTTCTCCTAATTTTCCCTTTGCC | 300 | no | 1013 | |
|  |  |  |  |  | GYPB-103k+277-j3R |  | CCGCTCCTGTGGCTTTGCCTA | 300 |  |  | |
| 103-5.3 |  |  | "short" for *GYPE* wildtypes |  | GYPE-103k-779-j53F |  | AAATGCCGCCCGAACCGA | 300 | no | 1583 | |
|  |  |  |  |  | GYPE-103k+762-j53R |  | CCACTACATTCCTCCCACAACGCA | 300 |  |  | |
| 103-5.4 |  |  | "short" for 103 kb *GYPB* deletion |  | GYPA-103k-440-j5F |  | CACAAACTCAGTAATTCTCTTTAGAGTGACGC | 300 | no | 770 | |
|  |  |  |  |  | GYPB-103k+277-j3R |  | CCGCTCCTGTGGCTTTGCCTA | 300 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 103-6.1 |  |  | "long" for *GYPA* wildtypes |  | GYPA-103k-1262-F |  | GGTTTTCATGCCTCTGGGATGGT | 300 | no | 2669 | |
|  |  |  |  |  | GYPA-103k+1356-R |  | AATAAGAATTTACTGCAACCTTAGCCCG | 300 |  |  | |
| 103-6.2 |  |  | "long" for *GYPB* wildtypes |  | GYPB-103k-1253-F |  | TGCCTCTGGGATTGGGATTAGTGA | 300 | no | 2648 | |
|  |  |  |  |  | GYPB-103k+1344-R |  | TGCAGCCTTAGCACATTGGATATTCTC | 300 |  |  | |
| 103-6.3 |  |  | "long" for *GYPE* wildtypes |  | GYPE-103k-1261-F |  | TAGGTTTTCATGCCTCTGGGATTTAG | 300 | no | 2834 | |
|  |  |  |  |  | GYPE-103k+1518-R |  | GTGTGATATAGGAAGAAGCAATCGGTGTA | 300 |  |  | |
| 103-6.4 |  |  | "long" for 103 kb *GYPB* deletion |  | GYPA-103k-1262-F |  | GGTTTTCATGCCTCTGGGATGGT | 300 | no | 2656 | |
|  |  |  |  |  | GYPB-103k+1344-R |  | TGCAGCCTTAGCACATTGGATATTCTC | 300 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 019-3.1 |  |  | *GYPA* wildtype breakpoint 5' |  | GYPA\_18k-1320j5F |  | GAGCCCATTTGGTATCCACAGTGC | 300 | no | 2200 | |
|  |  |  |  |  | GYPall\_18k+890j5R |  | TCACATCCCTCCAGTCCCTGAGC | 300 |  |  | |
| 019-3.2 |  |  | *GYPA* wildtype breakpoint 3' |  | GYPall\_18k-866j3F |  | TTCCCTGCTGCACCTTCCTTACC | 300 | no | 1400 | |
|  |  |  |  |  | GYPA-18k+549-j3R |  | GTTGTTCCCCTGAAAATGTTAACTCTTGA | 300 |  |  | |
| 019-3.3 |  |  | *GYPB* wildtype breakpoint 5' |  | GYPB\_18k-787j5F |  | GTTAAACTTCAAAATCCCTATCACCCGA | 300 | no | 1718 | |
|  |  |  |  |  | GYPall\_18k+890j5R |  | TCACATCCCTCCAGTCCCTGAGC | 300 |  |  | |
| 019-3.4 |  |  | *GYPB* wildtype breakpoint 3' |  | GYPall\_18k-866j3F |  | TTCCCTGCTGCACCTTCCTTACC | 300 | no | 1373 | |
|  |  |  |  |  | GYPB-18k+458-j3R |  | CTGATCTTAATTAATGGATACCAATGTAGGGTAGTC | 300 |  |  | |
| 019-3.5 |  |  | *GYPE* wildtype breakpoint 5' |  | GYPE\_18k-1237j5F |  | AGGAGTGGAGGATTAAAAGGAGAAGCA | 300 | no | 2110 | |
|  |  |  |  |  | GYPall\_18k+890j5R |  | TCACATCCCTCCAGTCCCTGAGC | 300 |  |  | |
| 019-3.6 |  |  | *GYPE* wildtype breakpoint 3' |  | GYPall\_18k-866j3F |  | TTCCCTGCTGCACCTTCCTTACC | 300 | no | 1300 | |
|  |  |  |  |  | GYPE-18k+448-j3R |  | TGGATACCAATGTAGGTACTTTGGATTGTG | 300 |  |  | |
| 019-3.7 |  |  | 19 kb *GYPB* deletion |  | GYPB-18k-766-j5F |  | CCTATCACCAGATTATTGTTTAAAAGCATTCA | 300 | no | 1292 | |
|  |  |  |  |  | GYPB-18k+458-j3R |  | CTGATCTTAATTAATGGATACCAATGTAGGGTAGTC | 300 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **B. Sequencing Primers** |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 110-3.5 |  |  | sequencing unit *GYPA* |  | GYPA-110k+599-j3R |  | ATGATTTGTGAGCACCTCCT | n.a | no | n.a. | |
| 110-3.6 |  |  | sequ. unit *GYPE* & 110 kb GYPB del. |  | GYPE-110k+401-j3R |  | GGGTTGAGGTTCTGGAAAGG | n.a | no | n.a. | |
| 110-3.7 |  |  | sequencing all amplicons |  | GYPABE-110k-825-F |  | TATCCCATGCAGGTTGTCAG | n.a | no | n.a. | |
| 110-3.8 |  |  | sequencing all amplicons |  | GYPABE-110k-298-F |  | GAGGATGTGTGTTCAGGAGC | n.a | no | n.a. | |
| 110-4.5 |  |  | sequ. all "long" amplicons |  | GYPABE-110k-248-R |  | TCAGGAGACACCCCAAATGC | n.a | no | n.a. | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 103-5.5 |  |  | sequencing unit *GYPA* |  | GYPA-103k+318-j5R |  | TGCTGATGCAAGAGGTGGGT | n.a | no | n.a. | |
| 103-5.6 |  |  | sequencing unit *GYPB* |  | GYPB-103k-466-j3F |  | GCCCCCGCAGTGAGCACAAG | n.a | no | n.a. | |
| 103-5.7 |  |  | sequencing unit *GYPE* |  | GYPE-103k-472-F |  | GATCTGGCCCCAGCAGTGAA | n.a | no | n.a. | |
| 103-5.8 |  |  | sequencing 103 kb *GYPB* deletion |  | GYPA-103k-384-j5F |  | GTGGCTAATTATGGAACAAC | n.a | no | n.a. | |
| 103-6.5 |  |  | sequ. all "long" amplicons |  | GYPABE-103k-219-F |  | ATCACTCAATATGTGGACTG | n.a | no | n.a. | |
| 103-6.6 |  |  | sequ. all "long" amplicons |  | GYPABE-103k+912-R |  | GCAGAAGGGGAAGCAAACAC | n.a | no | n.a. | |
| 103-6.7 |  |  | sequ. all "long" amplicons |  | GYPABE-103k-219-F |  | ATCACTCAATATGTGGACTG | n.a | no | n.a. | |
| 103-6.8 |  |  | sequ. all "long" amplicons |  | GYPABE-103k-219-F |  | ATCACTCAATATGTGGACTG | n.a | no | n.a. | |
| 103-6.9 |  |  | sequ. all "long" amplicons |  | GYPABE-103k+180-R |  | TCCTTATAGCAGCACCCGAC | n.a | no | n.a. | |
| 103-6.10 |  |  | sequ. all "long" amplicons |  | GYPABE-103k-219-F |  | ATCACTCAATATGTGGACTG | n.a | no | n.a. | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 019-4.1 |  |  | sequ. all breakpoint 5' amplicons |  | GYPABE-18k-548-F |  | CAATAGATCATGTCTTCGCC | n.a | no | n.a. | |
| 019-4.2 |  |  | sequ. all breakpoint 3' amplicons |  | GYPall\_18k-196j3F |  | GCTCACTACTGTCAGCCCTC | n.a | no | n.a. | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **C. Diagnostic Gap-PCRs & PCR-SSPs** |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 110-1.3 |  |  | 110 kb *GYPB* deletion |  | GYPB-110k-272-j5-F |  | TGGGTGAGCAGGGCTGAGATTC | 600 | GH1, or GH7 | | 917 | |
|  |  |  |  |  | GYPE-110k+179-j3R |  | CAATGTCTAGGATATCAATTCATTCTGCGA | 600 | GH1, or GH7 | |  | |
| 110-1.1 |  |  | 110 kb wildtype 5' |  | GYPB-110k-272-j5-F |  | TGGGTGAGCAGGGCTGAGATTC | 900 | GH1, or GH7 | | 778 | |
|  |  |  |  |  | GYPB-110k+31-j5R |  | AATGTAACACTCACAAACTAAGTGTATAACTCATCTG | 900 | GH1, or GH7 | |  | |
| 110-1.2 |  |  | 110 kb wildtype 3' |  | GYPAE-110k-114-j3F |  | CCTCCACACGCTGCTTTGTCTTTC | 900 | GH3, or GH7 | | 349 | |
|  |  |  |  |  | GYPE-110k+179-j3R |  | CAATGTCTAGGATATCAATTCATTCTGCGA | 900 | GH3, or GH7 | |  | |
|  |  |  |  |  |  |  |  |  |  | |  | |
|  |  |  |  |  |  |  |  |  |  | |  | |
| 103-1.3 |  |  | 103 kb *GYPB* deletion |  | GYPA-103k-97-j5F |  | GAATCATTGCGGTGGTTTCCCTTA | 600 | GH5, or GH7 | | 215 | |
|  |  |  |  |  | GYPB-103k+66-j3R |  | GTTTAGTTGATCTACAGTTTTGCATGGCTT | 600 | GH5, or GH7 | |  | |
| 103-1.1 |  |  | 103 kb wildtype 5' |  | GYPA-103k-97-j5F |  | GAATCATTGCGGTGGTTTCCCTTA | 200 | GH6, or GH7 | | 365 | |
|  |  |  |  |  | GYPA-103k+229-j5R |  | TGCTTTCACGGGCTGTTATCCAA | 200 | GH6, or GH7 | |  | |
| 103-1.2 |  |  | 103 kb wildtype 3' |  | GYPB-103k-215-j3F |  | AGAGAAATCACTCAATATGTGGAATGGCTA | 900 | GH5, or GH7 | | 339 | |
|  |  |  |  |  | GYPB-103k+66-j3R |  | GTTTAGTTGATCTACAGTTTTGCATGGCTT | 900 | GH5, or GH7 | |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 019-1.3 |  |  | 19 kb deletion of *GYPB* |  | GYPB\_19k-352\_j5F |  | AGCCCTTCCACCTCCACTCCAC | 300 | GH7 | 800 | |
|  |  |  |  |  | GYPB\_19k+397\_j3R |  | TCAGGAACAGCAATGTAAACAGGCTAGAT | 300 | GH7 |  | |
| 019-1.1 |  |  | 19 kb wildtype 5' |  | GYPB\_19k-352\_j5F |  | AGCCCTTCCACCTCCACTCCAC | 300 | GH7 | 773 | |
|  |  |  |  |  | GYPB\_19k+378\_j5R |  | ATTGCATGAACCCAGGAGCCC | 300 | GH7 |  | |
| 019-1.2 |  |  | 19 kb wildtype 3' |  | GYPB\_19k-331\_j3F |  | CAAATTGTTTTCAGAGAGGCTCCCTTC | 300 | GH7 | 775 | |
|  |  |  |  |  | GYPB\_19k+397\_j3R |  | TCAGGAACAGCAATGTAAACAGGCTAGAT | 300 | GH7 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| S |  |  | *GYPB\*03*, c.147C |  | rs7683365Ss\_A\_F |  | GAAACGATGGACAAGTTGTCGCA | 600 | GH1 | 138 | |
|  |  |  |  |  | rs7683365Ss\_R |  | AAAATATGATTAAGAAAAGGAAACCCGC | 600 | GH1 |  | |
| s |  |  | *GYPB\*04*, c.147T |  | rs7683365Ss\_G\_F |  | TGAAACGATGGACAAGTTGTCACG | 600 | GH1 | 139 | |
|  |  |  |  |  | rs7683365Ss\_R |  | AAAATATGATTAAGAAAAGGAAACCCGC | 600 | GH1 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| He+ |  |  | *GYPB\*06.01/06.02*, c.59T>G, c60A>G |  | GYPB-i1-152F |  | AATAAGGCCAATAATACAATACTTACCACGC | 200 | GH1 | 227 | |
|  |  |  |  |  | GYPB+59-HeR |  | GCATTGCCACACCAGAGGTACTCC | 200 | GH1 |  | |
| non He+ |  |  | *GYPB\*03/04*, c.59T, c.60A |  | GYPB-i1-152F |  | AATAAGGCCAATAATACAATACTTACCACGC | 300 | GH1 | 227 | |
|  |  |  |  |  | GYPB+59-wtR |  | GCATTGCCACCTCAGTGGTACCTA | 300 | GH1 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| U+w, NY, He(NY) |  |  | *GYPB\*03N.01/03N.02*, c.230C>T |  | GYPB-e5-208-230-mutF |  | GTGTTTGATGGCTGGTATTATTGGGAT | 600 | GH1 | 145 | |
|  |  |  |  |  | GYPB-in5+51R |  | AACTGAACTCAGAGGAATAAACCCTGCT | 600 | GH1 |  | |
| non U+w, NY, He(NY) |  |  | *GYPB\*03/04*, c.230C |  | GYPB-e5-230wtF |  | TGTGTGATGGCTGGTATTATTGGAAC | 500 | GH1 | 144 | |
|  |  |  |  |  | GYPB-in5+51R |  | AACTGAACTCAGAGGAATAAACCCTGCT | 500 | GH1 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| U+w, P2, He(P2) |  |  | *GYPB\*03N.03/03N.04*, intron 5, +5g>t |  | GYPB-in5+5mutF |  | AGTATTCGCCGACTGATAAAGGTGAT | 300 | GH1 | 99 | |
|  |  |  |  |  | GYPB-in5+51R |  | AACTGAACTCAGAGGAATAAACCCTGCT | 300 | GH1 |  | |
| non U+w, P2, He(P2) |  |  | *GYPB\*03/04*, intron 5, +5g |  | GYPB-in+5wtF |  | ACACTATTCGCCGACTGATAAAGGTGAG | 300 | GH1 | 101 | |
|  |  |  |  |  | GYPB-in5+51R |  | AACTGAACTCAGAGGAATAAACCCTGCT | 300 | GH1 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **D. "Privates"** |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 110 kb "-800A" |  |  | rs143076335 T |  | GYPB-865F |  | CCTAGGGATGTGGCTTCCTGAGGAC | 500 | GH6 | 109 | |
|  |  |  |  |  | GYPB-110-730R |  | TGGGAGCTGGGTGAGGCGTT | 500 | GH6 |  | |
| 110 kb "+36A" |  |  | no rs number, GRCh38.p12 |  | nGYPBE-443F |  | GGCAGAAATGGATGGCTAGGGAACGCA | 500 | GH6 | 536 | |
|  |  |  | genomic coordinate: 143913980 A |  | nGYPE-110+36R |  | TATTAAAATGTAACACTCACAAACTGTAAGTGTGTAACTT | 500 | GH6 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| 103 kb "-402T" |  |  | rs186872886 A |  | nGYPA-103-402F |  | GTCTCACAACATCCTGTAACATTGGCGTAT | 500 | GH6 | 130 | |
|  |  |  |  |  | nGYPA-334R |  | TCTATTCACTATTCTTCCACAGAGAGGGACTTG | 500 | GH6 |  | |
| 103 kb "+70C" |  |  | rs142534144 G |  | GYPB103+11F |  | GGTCATGATTGTAAGTATTCTAAGACCTGCCA | 500 | GH6 | 120 | |
|  |  |  |  |  | nGYPB-103+69R |  | TCTGTTCTCATGCTGCTATAAAAAACTGGGC | 500 | GH6 |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| **E. Positive Amplification Controls** |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
| GH1 |  |  | pos.ampl. cont. 434 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 70 | no | 434 | |
|  |  |  |  |  | GH1+274-R |  | CCACTCACGGATTTCTGTTGTGTTTC | 70 |  |  | |
| GH2 |  |  | pos.ampl. cont. 434 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 60 | no | 434 | |
|  |  |  |  |  | GH1+274-R |  | CCACTCACGGATTTCTGTTGTGTTTC | 60 |  |  | |
| GH3 |  |  | pos.ampl. cont. 434 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 50 | no | 434 | |
|  |  |  |  |  | GH1+274-R |  | CCACTCACGGATTTCTGTTGTGTTTC | 50 |  |  | |
| GH4 |  |  | pos.ampl. cont. 1000 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 80 | no | 1000 | |
|  |  |  |  |  | HGH+521R |  | TGTCGAACTTGCTGTAGGTCTGCTTGA | 80 |  |  | |
| GH5 |  |  | pos.ampl. cont. 1000 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 60 | no | 1000 | |
|  |  |  |  |  | HGH+521R |  | TGTCGAACTTGCTGTAGGTCTGCTTGA | 60 |  |  | |
| GH6 |  |  | pos.ampl. cont. 1000 bp |  | GH1+96-F |  | TGCCTTCCCAACCATTCCCTTA | 50 | no | 1000 | |
|  |  |  |  |  | HGH+521R |  | TGTCGAACTTGCTGTAGGTCTGCTTGA | 50 |  |  | |
| GH7 |  |  | pos.ampl. cont. 480 bp |  | GH1+50-F |  | TCCTGGCTTTTGGCCTGCTCTG | 70 | no | 480 | |
|  |  |  |  |  | GH1+274-R |  | CCACTCACGGATTTCTGTTGTGTTTC | 70 |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  | |

**Supplementary Figure s1. Genomic coordinates of homology analysis of the three GYP units (GRCh38.p12, Primary Assembly, NC\_000004.12)**

The three *GYPB* deletions of 110 kb (blue), 19 kb (gray) and 103 kb (red) are shown in Suppl Figure 1 along with the exact genomic coordinates for homologous and non homologous exons and introns. The 1000G Coriell sample, originally identified as representing a putative ~32 kb *GYPB* deletion (Coriell HG02970\*A1) turned out to be heterozygous for the two deletions of 110 kb and 103 kb, causing a compound overlap void of any *GYPB* sequences with a calculated length of 33.3 kb (not shown, genomic coordinates 143 991 719 to 144 025 065).



.

**Supplementary Results**

**Linkage exploration of the 110 kb and 103 kb *GYPB* deletions**

Linkage of different SNPs to *GYPB\*03* and *GYPB\*04* (rs7683365) and the 110 kb and 103 kb *GYPB* deletions is shown in the supplementary table s2.

On either *GYPA\*01/01*, or *GYPA\*02/02* homozygous samples, the 110 kb *GYPB* deletion was found 4 times in linkage with *GYPA\*01* (M) and 1 time with *GYPA\*02* (N), whereas the 103 kb *GYPB* deletion was found 7 times in linkage with *GYPA\*01* (M) and 3 times with *GYPA\*02* (N), respectively. Full linkage for both *GYPB* deletions, specifically to either *GYPA\*01*, or *GYPA\*02* may therefore be excluded formally.

Hybrid sequences analysed for the 110 kb and 103 kb *GYPB* deletions showed two and three “private” nucleotides respectively (Figure 2A and 2B), not found in either of the wildtype sequences. Testing all CEPH DNAs for four of these specific nucleotides by PCR-SSP (suppl. table s1D), showed full linkage of the “private” positions -402T (genomic coordinate 144095375, rs186872886, A) and +70C (143991649, rs142534144, G), relative to the breakpoint of the 103 kb *GYPB* deletion with the deletion itself. “Private” nucleotides -800A (144025051, rs143076335 T) and +36A (143913980, no rs) relative to the breakpoint of the 110 kb *GYPB* deletion however, were also identified in CEPH DNAs lacking this deletion (suppl. table s2).

Supplementary table s3 shows a population diversity analysis for the specific “private” nucleotides mentioned above plus hybrid-specific nucleotide -3T (143991649, rs557669971 A), specifically found in the hybrid sequence of the 103 kb *GYPB* deletion. Respective low frequency variants, indicative of the GYPB deletions 110 kb and 103 kb were only identified among the 1000G super populations of Africa (AFR, n = 661 individuals, 1322 chromosomes 4) and America (AMR, n = 347 individuals, 692 chromosomes 4). In comparison, none of the low frequency variants were identified among the super populations of East Asia (EAS), Europe (EUR), and South Asia (SAS) (suppl. table s3). Allele frequencies of -800A and -402T among AFR were 0.0983 and 0.0204 in comparison to 0.1146 and 0.0510 among CEPH DNAs, respectively (suppl. tables s2 and s3).

**Supplementary Table s2.** Linkage exploration of CEPH samples.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | rs7683365 |  | C and/or T (GYPB\*03 and/or GYPB\*04) | C or T and 103 kb | C and T neg., del 103 kb homo | C or T and del 110 kb | C and T neg., del 103 kb, 110 kb hetero |  | totals |  | Allele / haplotype frequncy |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| ***GYPA\*01/02*** | **g.144120567** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| *GYPA\*01* | rs7682260 CC |  | 28 | 4 | 1 | 3 | 1 |  | 37 |  |  |
| *GYPA\*01/02* | rs7682260 CT |  | 69 | 7 | 0 | 5 | 0 |  | 81 |  |  |
| *GYPA\*02* | rs7682260 TT |  | 35 | 3 | 0 | 1 | 0 |  | 39 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | totals |  | 132 | 14 | 1 | 9 | 1 |  | 157 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **110 kb "-800A"** | **g.144025051** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| unit *GYPB* | rs143076335 TT |  | n.t. | n.t. | n.t. | n.t. | n.t. |  |  |  |  |
| unit *GYPB* | rs143076335 TN |  | 23 | 3 | 0 | 9 | 1 |  | 36 |  | 0,1146 |
| unit *GYPB* | rs143076335 NN |  | 109 | 11 | 1 | 0 | 0 |  | 121 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | totals |  | 132 | 14 | 1 | 9 | 1 |  | 157 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **110 kb "+36A"** | **g.143913980** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| unit *GYPE* | no rs, AA |  | n.t. | n.t. | n.t. | n.t. | n.t. |  |  |  |  |
| unit *GYPE* | no rs, AN |  | 10 | 11 | 1 | 9 | 1 |  | 32 |  | 0,1019 |
| unit *GYPE* | no rs, NN |  | 122 | 3 | 0 | 0 | 0 |  | 125 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | totals |  | 132 | 14 | 1 | 9 | 1 |  | 157 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **103 kb "-402T"** | **g.144095375** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| unit *GYPA* | rs186872886 AA |  | n.t. | n.t. | n.t. | n.t. | n.t. |  |  |  |  |
| unit *GYPA* | rs186872886 AN |  | 0 | 14 | 1 | 0 | 1 |  | 16 |  | 0,0510 |
| unit *GYPA* | rs186872886 NN |  | 132 | 0 | 0 | 9 | 0 |  | 141 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | totals |  | 132 | 14 | 1 | 9 | 1 |  | 157 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **103 kb "-3T"** | **g.143991649** |  |  |  |  |  |  |  |  |  |  |
| unit *GYPA* | rs557669971 A |  | n.t. |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **103 kb "+70C"** | **g.143991649** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| unit *GYPB* | rs142534144 GG |  | n.t. | n.t. | n.t. | n.t. | n.t. |  |  |  |  |
| unit *GYPB* | rs142534144 GN |  | 0 | 14 | 1 | 0 | 1 |  | 16 |  | 0,0510 |
| unit *GYPB* | rs142534144 NN |  | 132 | 0 | 0 | 9 | 0 |  | 141 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | totals |  | 132 | 14 | 1 | 9 | 1 |  | 157 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Supplementary Table s3.** “Privates” SNP variation in the 1000G super populations.

East Asia (EAS, n chromosomes 4 = 1.008), Europe (EUR, n = 1.006), Africa (AFR, n = 1.322), America (AMR, n = 694) and South Asia (SAS, n = 978)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | |  |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| "private nucleotides" | genomic | rs number |  | EAS | | | EUR | | AFR | | AMR | | SAS | |
| relative to breakpoint | coordinates | nucleotide |  | 1008 | | | 1006 | | 1322 | | 694 | | 978 | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| 110 kb "-800A" | g.144025051 | rs143076335 T |  | 0 | | | 0 | | 130 | | 4 | | 0 | |
|  |  | allele-frequency: |  |  | | |  | | 0,0983 | | 0,0058 | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| 110 kb "+36A" | g.143913980 | n.a. |  | n.a. | | | n.a. | | n.a. | | n.a. | | n.a. | |
|  |  |  |  |  | | |  | | n.a. | | n.a. | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| 103 kb "-402T" | g.144095375 | rs186872886 A |  | 0 | | | 0 | | 27 | | 1 | | 0 | |
|  |  | allele-frequency: |  |  | | |  | | 0,0204 | | 0,0014 | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| 103 kb "-3T" | g.143991649 | rs557669971 A |  | 0 | | | 0 | | 4 | | 1 | | 0 | |
|  |  | allele-frequency: |  |  | | |  | | 0,0030 | | 0,0014 | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
| 103 kb "+70C" | g.143991649 | rs142534144 G |  | 0 | | | 0 | | 28 | | 1 | | 0 | |
|  |  | allele-frequency: |  |  | | |  | | 0,0212 | | 0,0014 | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  |  | | |  | |  | |  | |  | |
|  |  |  |  | |  |  | |  | |  | |  | |

.

**Supplementary Table s4. CEPH samples, individual results and genotypes deduced.**

Abbreviations: African region (Region): Subsaharan Afrjca (sub), North Africa (north) Nationalities: Central African Republic (CAR), Democratic Republic of Congo (DR Congo), Kenya (Kenya), South Africa (SAfrica), Senegal (Senegal), Algeria (Algeria), Namibia (Namibia), Nigeria (Nigeria). Coloured genotypes: GT underlayed in grey are assumed but could also have genotype GG. GG underlayed in grey are hemizygous due to the simultaneous presence of a *GYPB* deletion on the second parental chromosome. Homozygosity for TT and CC underlayed in grey was assumed. Two samples among Mandenka people from Senegal did not deliver results.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **method** |  | MALDI | MALDI | MALDI | MALDI |  | SSP | SSP | SSP |  | SSP | SSP | SSP |  | SSP | SSP | SSP |  | SSP |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **gene** |  | ABO | GYPA | on GYPB | GYPB |  | on GYPB | GYPB | GYPB |  | GYPB to  GYPE | GYPB | GYPE |  | GYPA to  GYPB | GYPA | GYPB |  | GYPB |  |  |  |
|  |  |  |  |  | **rs number** |  | rs8176719 | rs7682260 | rs7682260 | rs7683365 |  | rs7682260 | rs79492560 | rs139511876 |  | no rs | rs143076335 T | pos. 143913980 |  | no rs | rs186872886 A | rs142534144 G |  | no rs |  |  |  |
|  |  |  |  |  | **polymorphism** |  | 01 [d]|non01 [G] | GYPA\*01 [C]|*\*02* [T] | GYPB\*06 [G] | GYPB\*03 [T]|*\*04* [C] |  | GYPB\*06 [G] | GYPB [C]|GYPB\*03N.01/02 [T] | GYPB [G]|GYPB\*03N.03/04 [T] |  | 110 kb deletion GYPB | 110 kb "-800T>A" | 110 kb "+36G>A" |  | 103 kb deletion GYPB | 103 kb "-402C>T" | 103 kb "+70G>C" |  | 19 kb deletion GYPB |  |  |  |
|  |  |  |  |  | **ISBT** |  | c.261delG | c.59C>T | c.59T>G | c.143C>T |  | c.59T>G | c.230C>T | c.270+5g>t |  | n.a | n.a. | n.a. |  | n.a | n.a. | n.a. |  | n.a |  | 1st GYPB\* | 2nd GYPB\* |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CEPH ID | PAT-ID | Region | Nationality | Population | **Gender** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| HGDP00981 | 682865 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00452 | 681916 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00453 | 681917 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00454 | 681918 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00455 | 681919 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00457 | 681920 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00458 | 681921 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP00459 | 681922 | sub | CAR | Biaka\_Pygmy | **M** |  | GG | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00460 | 681923 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00461 | 681924 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00464 | 681925 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00465 | 681926 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | TT | G | TT |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03 | 03N.04 |
| HGDP00466 | 681927 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00469 | 681928 | sub | CAR | Biaka\_Pygmy | **M** |  | GG | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00470 | 681929 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00472 | 681930 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00473 | 681931 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00475 | 681932 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00479 | 681933 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | TT | neg | TC |  | TT | CC | GT |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.03 |
| HGDP00985 | 681934 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00986 | 681935 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | TC |  | TT | CC | GT |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 04 | 03N.03 |
| HGDP01087 | 682866 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP01086 | 681936 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | TC |  | TT | CC | GT |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.03 |
| HGDP01090 | 681937 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01094 | 681938 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01092 | 682867 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00448 | 682879 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | n.a. |  | TT | CC | GG |  | neg | TT | GA |  | pos (homo) | TT | CC |  | neg |  | del103 | del103 |
| HGDP01084 | 682880 | sub | CAR | Biaka\_Pygmy | **F** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01085 | 682881 | sub | CAR | Biaka\_Pygmy | **F** |  | GG | TT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03 | 03N.04 |
| HGDP01091 | 682882 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00980 | 682898 | sub | CAR | Biaka\_Pygmy | **F** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00451 | 682900 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00477 | 682901 | sub | CAR | Biaka\_Pygmy | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP01088 | 682902 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP01089 | 682903 | sub | CAR | Biaka\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01093 | 682904 | sub | CAR | Biaka\_Pygmy | **M** |  | GG | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00449 | 681939 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | Gd | CT | G | TT |  | GG | CC | GT |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 03N.04 | del110 |
| HGDP00450 | 681940 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | Gd | CT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03N.04 | 03N.04 |
| HGDP00456 | 681941 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | TT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03N.04 | 03N.04 |
| HGDP00462 | 681942 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | CT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03N.04 | 03N.04 |
| HGDP00463 | 681943 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00467 | 681944 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP00471 | 681945 | sub | DR Congo | Mbuti\_Pygmy | **F** |  | GG | CT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP00474 | 681946 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | TT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | pos | CT | GC |  | neg |  | 03N.04 | del103 |
| HGDP00476 | 681947 | sub | DR Congo | Mbuti\_Pygmy | **F** |  | Gd | CT | G | TT |  | GG | CC | GT |  | neg | TA | GA |  | pos | CT | GC |  | neg |  | 03N.04 | del103 |
| HGDP00478 | 681948 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | dd | CC | neg | n.a. |  | TT | CC | GG |  | pos | TA | GA |  | pos | CT | GC |  | neg |  | del110 | del103 |
| HGDP00982 | 681949 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | GG | CT | G | CC |  | GG | CC | GG |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | 06N.01 | 06N.01 |
| HGDP00984 | 681950 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | Gd | CT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP01081 | 681951 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | Gd | CT | G | TT |  | GG | CC | GT |  | neg | TA | GG |  | pos | CT | GC |  | neg |  | 03N.04 | del103 |
| HGDP00468 | 682905 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00983 | 682906 | sub | DR Congo | Mbuti\_Pygmy | **M** |  | GG | CT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP01405 | 681966 | sub | Kenya | Bantu\_N.E. | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01406 | 681967 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01408 | 681968 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01411 | 681969 | sub | Kenya | Bantu\_N.E. | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01412 | 681970 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01414 | 681971 | sub | Kenya | Bantu\_N.E. | **F** |  | GG | CT | neg | TC |  | GT | CT | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.02 |
| HGDP01415 | 681972 | sub | Kenya | Bantu\_N.E. | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01416 | 681973 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01417 | 681974 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CT | neg | TT |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 03 |
| HGDP01418 | 681975 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CT | neg | TC |  | GT | CT | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.02 |
| HGDP01419 | 681976 | sub | Kenya | Bantu\_N.E. | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01413 | 682908 | sub | Kenya | Bantu\_N.E. | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00993 | 681958 | sub | SAfrica | Bantu\_S.E.\_Pedi | **M** |  | GG | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00994 | 681959 | sub | SAfrica | Bantu\_S.E.\_S.Sotho | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01028 | 681960 | sub | SAfrica | Bantu\_S.W.\_Herero | **M** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01030 | 681961 | sub | SAfrica | Bantu\_S.E.\_Tswana | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01031 | 681962 | sub | SAfrica | Bantu\_S.W.\_Ovambo | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01033 | 681963 | sub | SAfrica | Bantu\_S.E.\_Zulu | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01034 | 681964 | sub | SAfrica | Bantu\_S.E.\_Tswana | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01035 | 681965 | sub | SAfrica | Bantu\_S.W.\_Herero | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP00904 | 681999 | sub | Senegal | Mandenka | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00905 | 682000 | sub | Senegal | Mandenka | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00906 | 682001 | sub | Senegal | Mandenka | **M** |  | dd | CT | G | n.a. |  | GG | CC | GG |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | n.a. | n.a. |
| HGDP00907 | 682002 | sub | Senegal | Mandenka | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00908 | 682003 | sub | Senegal | Mandenka | **M** |  | GG | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00909 | 682004 | sub | Senegal | Mandenka | **F** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00910 | 682005 | sub | Senegal | Mandenka | **F** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00911 | 682006 | sub | Senegal | Mandenka | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00912 | 682007 | sub | Senegal | Mandenka | **M** |  | Gd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00913 | 682008 | sub | Senegal | Mandenka | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00914 | 682009 | sub | Senegal | Mandenka | **F** |  | dd | CC | neg | n.a. |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | n.a. | n.a. |
| HGDP00915 | 682010 | sub | Senegal | Mandenka | **F** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00917 | 682011 | sub | Senegal | Mandenka | **F** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00918 | 682012 | sub | Senegal | Mandenka | **F** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01199 | 682013 | sub | Senegal | Mandenka | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP01200 | 682014 | sub | Senegal | Mandenka | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01201 | 682015 | sub | Senegal | Mandenka | **F** |  | dd | CT | G | TC |  | GG | CC | GG |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 03 | 06N.01 |
| HGDP01202 | 682016 | sub | Senegal | Mandenka | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01283 | 682017 | sub | Senegal | Mandenka | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01284 | 682018 | sub | Senegal | Mandenka | **M** |  | Gd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01285 | 682019 | sub | Senegal | Mandenka | **M** |  | Gd | CT | G | TC |  | GT | CC | GT |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP01286 | 682020 | sub | Senegal | Mandenka | **M** |  | dd | TT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00916 | 682883 | sub | Senegal | Mandenka | **F** |  | Gd | TT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP00919 | 682884 | sub | Senegal | Mandenka | **M** |  | GG | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01253 | 682744 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01254 | 682745 | north | Algeria | Mozabite | **F** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01255 | 682746 | north | Algeria | Mozabite | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP01256 | 682747 | north | Algeria | Mozabite | **M** |  | dd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01257 | 682748 | north | Algeria | Mozabite | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01258 | 682749 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01259 | 682750 | north | Algeria | Mozabite | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01260 | 682751 | north | Algeria | Mozabite | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01261 | 682752 | north | Algeria | Mozabite | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP01262 | 682753 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | TT |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 03 |
| HGDP01263 | 682754 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01264 | 682755 | north | Algeria | Mozabite | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01265 | 682756 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01266 | 682757 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01267 | 682758 | north | Algeria | Mozabite | **F** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01268 | 682759 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01269 | 682760 | north | Algeria | Mozabite | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01270 | 682761 | north | Algeria | Mozabite | **F** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01271 | 682762 | north | Algeria | Mozabite | **M** |  | dd | TT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01272 | 682763 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01273 | 682764 | north | Algeria | Mozabite | **F** |  | Gd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01274 | 682765 | north | Algeria | Mozabite | **F** |  | dd | CT | G | TC |  | GT | CC | GT |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP01275 | 682766 | north | Algeria | Mozabite | **F** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01276 | 682767 | north | Algeria | Mozabite | **F** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01277 | 682768 | north | Algeria | Mozabite | **F** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01278 | 682769 | north | Algeria | Mozabite | **M** |  | dd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01279 | 682770 | north | Algeria | Mozabite | **M** |  | Gd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01280 | 682771 | north | Algeria | Mozabite | **F** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01282 | 682772 | north | Algeria | Mozabite | **M** |  | dd | CT | neg | TT |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 03 | del110 |
| HGDP01281 | 682928 | north | Algeria | Mozabite | **F** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00987 | 681952 | sub | Namibia | San | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00991 | 681953 | sub | Namibia | San | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00992 | 681954 | sub | Namibia | San | **M** |  | dd | CC | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP01029 | 681955 | sub | Namibia | San | **M** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01032 | 681956 | sub | Namibia | San | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP01036 | 681957 | sub | Namibia | San | **M** |  | Gd | CT | G | TC |  | GG | CC | GG |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | 03 | 06N.01 |
| HGDP00988 | 682907 | sub | Namibia | San | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00920 | 681977 | sub | Nigeria | Yoruba | **F** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00924 | 681978 | sub | Nigeria | Yoruba | **F** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00925 | 681979 | sub | Nigeria | Yoruba | **F** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00926 | 681980 | sub | Nigeria | Yoruba | **F** |  | dd | CT | G | TC |  | GT | CC | GT |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP00927 | 681981 | sub | Nigeria | Yoruba | **M** |  | dd | CT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00928 | 681982 | sub | Nigeria | Yoruba | **F** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00929 | 681983 | sub | Nigeria | Yoruba | **M** |  | dd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00930 | 681984 | sub | Nigeria | Yoruba | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00931 | 681985 | sub | Nigeria | Yoruba | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00932 | 681986 | sub | Nigeria | Yoruba | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00933 | 681987 | sub | Nigeria | Yoruba | **F** |  | Gd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00934 | 681988 | sub | Nigeria | Yoruba | **F** |  | dd | CC | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP00935 | 681989 | sub | Nigeria | Yoruba | **F** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00936 | 681990 | sub | Nigeria | Yoruba | **M** |  | Gd | CT | neg | TT |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 03 | del103 |
| HGDP00937 | 681991 | sub | Nigeria | Yoruba | **M** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00938 | 681992 | sub | Nigeria | Yoruba | **F** |  | Gd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00939 | 681993 | sub | Nigeria | Yoruba | **F** |  | dd | CT | neg | CC |  | TT | CC | GG |  | pos | TA | GA |  | neg | CC | GG |  | neg |  | 04 | del110 |
| HGDP00940 | 681994 | sub | Nigeria | Yoruba | **M** |  | dd | TT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00941 | 681995 | sub | Nigeria | Yoruba | **M** |  | Gd | TT | neg | TC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00942 | 681996 | sub | Nigeria | Yoruba | **M** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00943 | 681997 | sub | Nigeria | Yoruba | **M** |  | dd | TT | neg | TC |  | TT | CC | GG |  | neg | TA | GG |  | neg | CC | GG |  | neg |  | 03 | 04 |
| HGDP00944 | 681998 | sub | Nigeria | Yoruba | **M** |  | Gd | TT | G | TC |  | GT | CC | GT |  | neg | TA | GA |  | neg | CC | GG |  | neg |  | 04 | 03N.04 |
| HGDP00921 | 682909 | sub | Nigeria | Yoruba | **F** |  | dd | CT | neg | CC |  | TT | CC | GG |  | neg | TT | GA |  | pos | CT | GC |  | neg |  | 04 | del103 |
| HGDP00922 | 682910 | sub | Nigeria | Yoruba | **F** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |
| HGDP00923 | 682911 | sub | Nigeria | Yoruba | **M** |  | Gd | CC | neg | CC |  | TT | CC | GG |  | neg | TT | GG |  | neg | CC | GG |  | neg |  | 04 | 04 |

.

**Supplementary Table s5. Accession number summary of GenBank submission**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| **Sequence Description** | **MNSs phenotypes of alleles/samples** | **Sequence ID** | **GenBank accession number** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| breakp. of *GYPB* del 19 kb, allele ***GYPB\*05N.03*** | S-s-U- | BP\_GYPBdel19kb\_HG01880\*C1 | MN005662 |
| breakp. of *GYPB* del 103 kb, allele ***GYPB\*05N.02*** | S-s-U- | BP\_GYPB-del103kb\_Lorena | MN005663 |
| breakp. of *GYPB* del 110 kb, allele ***GYPB\*05N.01*** | S-s-U- | BP\_GYPB-del110kb\_Gerold | MN005664 |
|  |  |  |  |
|  |  |  |  |
| wildtype at breakp. 19 kb, unit *GYPA* | MM / SS | GYPA\_del19kb-wt\_Verena | MN005665 |
| wildtype at breakp. 19 kb, unit *GYPA* | MM / ss | GYPA\_del19kb-wt\_Bruno | MN005666 |
| wildtype at breakp. 19 kb, unit *GYPA* | NN / SS | GYPA\_del19kb-wt\_Helmut | MN005667 |
| wildtype at breakp. 19 kb, unit *GYPA* | NN / ss | GYPA\_del19kb-wt\_Urs | MN005668 |
|  |  |  |  |
| wildtype at breakp. 19 kb, unit *GYPB* | MM / SS | GYPB\_del19kb-wt\_Verena | MN005669 |
| wildtype at breakp. 19 kb, unit *GYPB* | MM / ss | GYPB\_del19kb-wt\_Bruno | MN005670 |
| wildtype at breakp. 19 kb, unit *GYPB* | NN / SS | GYPB\_del19kb-wt\_Helmut | MN005671 |
| wildtype at breakp. 19 kb, unit *GYPB* | NN / ss | GYPB\_del19kb-wt\_Urs | MN005672 |
|  |  |  |  |
| wildtype at breakp. 19 kb, unit *GYPE* | MM / SS | GYPE\_del19kb-wt\_Verena | MN005673 |
| wildtype at breakp. 19 kb, unit *GYPE* | MM / ss | GYPE\_del19kb-wt\_Bruno | MN005674 |
| wildtype at breakp. 19 kb, unit *GYPE* | NN / SS | GYPE\_del19kb-wt\_Helmut | MN005675 |
| wildtype at breakp. 19 kb, unit *GYPE* | NN / ss | GYPE\_del19kb-wt\_Urs | MN005676 |
|  |  |  |  |
|  |  |  |  |
| wildtype at breakp. 103 kb, unit *GYPA* | MM / SS | GYPA\_del103kb-wt\_Verena | MN005677 |
| wildtype at breakp. 103 kb, unit *GYPA* | MM / ss | GYPA\_del103kb-wt\_Bruno | MN005678 |
| wildtype at breakp. 103 kb, unit *GYPA* | NN / SS | GYPA\_del103kb-wt\_Helmut | MN005679 |
| wildtype at breakp. 103 kb, unit *GYPA* | NN / ss | GYPA\_del103kb-wt\_Urs | MN005680 |
|  |  |  |  |
| wildtype at breakp. 103 kb, unit *GYPB* | MM / SS | GYPB\_del103kb-wt\_Verena | MN005681 |
| wildtype at breakp. 103 kb, unit *GYPB* | MM / ss | GYPB\_del103kb-wt\_Bruno | MN005682 |
| wildtype at breakp. 103 kb, unit *GYPB* | NN / SS | GYPB\_del103kb-wt\_Helmut | MN005683 |
| wildtype at breakp. 103 kb, unit *GYPB* | NN / ss | GYPB\_del103kb-wt\_Urs | MN005684 |
|  |  |  |  |
| wildtype at breakp. 103 kb, unit *GYPE* | MM / SS | GYPE\_del103kb-wt\_Verena | MN005685 |
| wildtype at breakp. 103 kb, unit *GYPE* | NN / SS | GYPE\_del103kb-wt\_Helmut | MN005686 |
|  |  |  |  |
|  |  |  |  |
| wildtype at breakp. 110 kb, unit *GYPA* | MM / SS | GYPA\_del110kb-wt\_Verena | MN005687 |
| wildtype at breakp. 110 kb, unit *GYPA* | MM / ss | GYPA\_del110kb-wt\_Bruno | MN005688 |
| wildtype at breakp. 110 kb, unit *GYPA* | NN / SS | GYPA\_del110kb-wt\_Helmut | MN005689 |
| wildtype at breakp. 110 kb, unit *GYPA* | NN / ss | GYPA\_del110kb-wt\_Urs | MN005690 |
|  |  |  |  |
| wildtype at breakp. 110 kb, unit *GYPB* | MM / SS | GYPB\_del110kb-wt\_Verena | MN005691 |
| wildtype at breakp. 110 kb, unit *GYPB* | MM / ss | GYPB\_del110kb-wt\_Bruno | MN005692 |
| wildtype at breakp. 110 kb, unit *GYPB* | NN / SS | GYPB\_del110kb-wt\_Helmut | MN005693 |
| wildtype at breakp. 110 kb, unit *GYPB* | NN / ss | GYPB\_del110kb-wt\_Urs | MN005694 |
|  |  |  |  |
| wildtype at breakp. 110 kb, unit *GYPE* | MM / SS | GYPE\_del110kb-wt\_Verena | MN005695 |
| wildtype at breakp. 110 kb, unit *GYPE* | MM / ss | GYPE\_del110kb-wt\_Bruno | MN005696 |
| wildtype at breakp. 110 kb, unit *GYPE* | NN / SS | GYPE\_del110kb-wt\_Helmut | MN005697 |
| wildtype at breakp. 110 kb, unit *GYPE* | NN / ss | GYPE\_del110kb-wt\_Urs | MN005698 |
|  |  |  |  |

**Supplementary Table s6. Homology analysis of the three *GYP* units on single nucleotide level**

Homology analysis of the three *GYP* units was done on the single nucleotide level across the complete *GYP* locus. The resulting alignment sequence for unit *GYPA* started with 2189 N, followed by the start of the homology region of unit *GYPA*, located at genomic coordinate 144147912 of NC\_000004.12 (GRCh38.p12, Primary Assembly), with a sequence reading GAAAAGCTACTCCTGCTTGC. Respective sequence had a final length of 293 bp, before entering 3N re-established alignment to unit *GYPB* and *GYPE*. Sequence of *GYPA* continued with the next starting sequence ATAGGTCCTGTAATATTTAC for 747 bp. Remark: Ns are continuously entered to align for positions of insertions, located on the other units. Alignment sequences for unit *GYPB* and *GYPE* were built up similarly. Alignments were done using Software GeneRunner, version 5.1.0.6. Beta [s5-s7].

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **unit** |  |  | **unit** |  |  | **unit** |  |  |
| ***GYPA*** |  | *.* | ***GYPB*** |  | *.* | ***GYPE*** |  | . |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| NNNNNNNNNNNNNNNNNNNN | 2189 | . | TGAGAGGCAATTATCTGAGT | 1548 | . | TGAGAGGCAATTATTTGAGT | 290 | . |
| GAAAAGCTACTCCTGCTTGC | 293 | . | N | 1 | . | N | 1 | . |
| NNN | 3 | . | TTTTTCACTGTTGGTGGGAC | 639 | . | GAAAAAAAGGAGAAGGAATT | 3018 | . |
| ATAGGTCCTGTAATATTTAC | 747 | . | N | 1 | . | NNNNNNNNN | 9 | . |
| N | 1 | . | GAAAAGCTACTCCTGCTTGC | 293 | . | TTTTCACTTCCTTAAATTGG | 1201 | . |
| CCTGAAGTTTTGATTGTT | 18 | . | NNN | 3 | . | NNNN | 4 | . |
| NNNN | 4 | . | ATAGGTCCTGTAATATTTAC | 747 | . | ATATGTGTGTATATATATAT | 22 | . |
| ATTTATGCTATCTATTTAAT | 1638 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | CCTGAAGTTTTGATTGTTTG | 76 | . | GAGTTACACACTTACAGTTT | 346 | . |
| CATTGCTTGCTCTGGAAAAC | 137 | . | NNNNNNNN | 8 | . | N | 1 | . |
| NNNN | 4 | . | TTTTTCACTTCCTTAAATTG | 1244 | . | CATTGCTTGCTCTGGAAAAC | 754 | . |
| CCCAATTAAGCTGTCAGTTC | 605 | . | NN | 2 | . | NNNNNN | 6 | . |
| N | 1 | . | GTTTGTGAGTGTTACATTTT | 338 | . | GTCTGAGAGGAAGCATTAAA | 422 | . |
| TAACACATACCATGTCTGAG | 402 | . | NNNN | 4 | . | N | 1 | . |
| N | 1 | . | CTGGAAAACCTTTCCAGAAC | 735 | . | GCTAAATCCATACGTTTAAA | 51 | . |
| AATTAAAATATACCTTCTTT | 138 | . | N | 1 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 3545 | . | TAACACATACCATGTCTGAG | 541 | . | ACTCATAATCTCTTTTCCAT | 3653 | . |
| TAATATCTGTTCTGTCTCTC | 87 | . | NNNNNNNNNNNNNNNNNNNN | 3545 | . | N | 1 | . |
| N | 1 | . | TAATATCTGTTCTGCCTCTC | 87 | . | TTGTAGTTGTTTTTTTATTT | 1601 | . |
| GAAATGATAAACTAACAAAG | 1320 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | CAAATGATAAACAAACAAAG | 484 | . | GAAAAAAAGAATCCAATAGG | 539 | . |
| TCATAGGGTGTGCCCACTTA | 206 | . | NN | 2 | . | NNN | 3 | . |
| NN | 2 | . | GAGTCAAAATTCTTAAAAGA | 1960 | . | TCCCTCAATCTCAGGCTCCT | 1517 | . |
| GAGTAGGTGTATTCCAGGTG | 37 | . | NNNNNNN | 7 | . | NNNNNNN | 7 | . |
| N | 1 | . | TGTCCTCACATTTCCATACC | 1320 | . | ATATGGCTTTTAGTTTTAAA | 211 | . |
| GAAAAAAAGAATCCAATAGG | 539 | . | N | 1 | . | NNNN | 4 | . |
| NNN | 3 | . | AAAAACAAAAACAACTCTTA | 73 | . | TATCATAGGGTCATTATTAA | 28 | . |
| TCCCTCATTCTCAGGCTCCT | 65 | . | NNNN | 4 | . | NNNN | 4 | . |
| N | 1 | . | TATCATAGGGTCATTATTAA | 292 | . | CAATAATTTGAATTTAACTT | 260 | . |
| GTAAATAATTTTTTTTCTCT | 1595 | . | NNNNNNNNNNNNNNNNNNNN | 39 | . | NNNNNNNNNNNNNNNNNNNN | 26 | . |
| N | 1 | . | ATCAGCATTTTCTAAGATAC | 411 | . | ACAGGGAAGAGAGATCAGCA | 345 | . |
| AAAAACAAAAACAACTCTTA | 369 | . | N | 1 | . | NNNN | 4 | . |
| NNNNNNNNNNNNNNNNNNNN | 39 | . | GAGTCAGAGAGACAAGAAAA | 102 | . | TTATTAAAATAAGCAGTAGC | 178 | . |
| ATCAGCATTTTCTAAGATAC | 332 | . | NNNNN | 5 | . | NNN | 3 | . |
| NNNN | 4 | . | CAAATGAAAGGGATGTAGTG | 72 | . | AACAAATGAAAGGGATGTAG | 542 | . |
| TTATTAAAATAAGCAGTAGC | 255 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GATGTATTCTTTGTGTTGCT | 84 | . | GACGATTGCCTTGCTTTCTC | 92 | . |
| GATGTATTCTTTGTGTTGCT | 84 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TGTGCCCAGCATTTGGGAAA | 219 | . | ATTTTTTCTAATATGACTTG | 816 | . |
| TGTGCCCAGCATTTGGGAAA | 141 | . | NNN | 3 | . | NNNNNNN | 7 | . |
| NNNNNNNN | 8 | . | AGAAACCCTCCAAGGAAACT | 160 | . | GAAAAAGGAAGGAAAGAAGA | 263 | . |
| GTCACTGGGGCAAGATGACC | 201 | . | N | 1 | . | NNNNNNNNNNNN | 12 | . |
| NNNNNNNNNNNNN | 13 | . | GACGATTGCCTTGCTTTCTC | 444 | . | CAGGGATCCAAGAAGAGAGC | 1284 | . |
| TATGTTGCTATAAAAAAAAA | 464 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TGGAGCCAACCCTGAAAATT | 220 | . | TAGCACTCACTTTGTGCCAT | 80 | . |
| CGGAGCCAACCCTGAAAATT | 220 | . | N | 1 | . | NNNNN | 5 | . |
| N | 1 | . | GTGTTTACTTAAAAATTCAT | 350 | . | GTTTACCATTTGAACCATTT | 855 | . |
| GTGTTTACTTAAAAATTCAT | 243 | . | NNNNN | 5 | . | NNNNNNNNNNNNNNNNNNNN | 804 | . |
| NNNNNN | 6 | . | AGGAAATATTTGTTTTCATC | 866 | . | TTTAAATCTATTTTTGTCTT | 177 | . |
| GAAAAAGGGAAGGAAAGAAG | 262 | . | NN | 2 | . | NN | 2 | . |
| NNNNNNNNNNNNNN | 14 | . | CCCCAATCATACTTATATGG | 169 | . | TTTTTT | 6 | . |
| CAGGGATCCAAGAAGAGAGC | 31 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | TTCCGGGCTTATCTCTGAGT | 416 | . | CTTTTGTTGCTTACGCTTTT | 490 | . |
| CTATTTAATTGGCACAGGAA | 599 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TAGCACTTACTTTGTGCCAT | 357 | . | AATTATGTGCATTTTAACTT | 800 | . |
| AAGTTATCTTTATAGATTAT | 64 | . | NNNNNNNNN | 9 | . | N | 1 | . |
| NN | 2 | . | AGTTCATCCATATTGTAGCG | 1563 | . | GCTTTCTTGATCTATCACAA | 681 | . |
| ACCCAATCCTACTTACATGG | 169 | . | NN | 2 | . | NNNNN | 5 | . |
| N | 1 | . | CTTTTGTTGCTTACGCTTTT | 490 | . | GATTCCTTGTTTCTGCTCTT | 178 | . |
| TTCCAGGTTTATCTCTGAGT | 39 | . | NNN | 3 | . | N | 1 | . |
| NNNNNN | 6 | . | TTATGCGCATTTTAACTTAA | 798 | . | ACTTTCTGGAGCTAGATAAG | 6474 | . |
| TTTTTCTCATTTTAAATACT | 786 | . | N | 1 | . | NNNNNNNNNNNNN | 13 | . |
| NNNNNNNNNNNNNN | 14 | . | GCTTTCTTGATCTATCACAA | 480 | . | CAAAATAATAATAATAACAT | 499 | . |
| TGTGTTGGTATATCTCATTT | 512 | . | N | 1 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNNNNN | 804 | . | TGAGCTCAGGCAATCCACCT | 383 | . | TGGCTCTAAAACAAAACCCT | 247 | . |
| TTTCAATCTATTTTTCTTTT | 597 | . | N | 1 | . | N | 1 | . |
| NNN | 3 | . | ACTTTCTGGAGCTAGATAAG | 295 | . | GATATCACAGATGATTCTAC | 66 | . |
| AAGATTGTTTTGCCTATTTA | 811 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | ATTTTATACAACTTAATTTT | 328 | . | GTATTGTTCAACCAGAGAAA | 20 | . |
| TCATATAAATCAATACCTTG | 748 | . | NNNNNNNNNNNNNNNNNNNN | 5285 | . | N | 1 | . |
| NNNNN | 5 | . | AGTGCCAGTGACAACCCACT | 596 | . | GGAAGTACCAGTGGTGTGAG | 125 | . |
| GATTCCTTGTTTCTGCTCTT | 803 | . | NN | 2 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 5285 | . | GTGAATATTATATGAAATTC | 728 | . | ACCTTAGACTAAAGCGGACA | 1065 | . |
| AGTGCCAGTGACAACCCACT | 566 | . | N | 1 | . | N | 1 | . |
| NNNNNNNNNNN | 11 | . | GATATCACAGATGATTCTAC | 66 | . | CTCCACCTCTGCTGTCTCAG | 2097 | . |
| TCAAAATAATAATAATAATA | 500 | . | N | 1 | . | NN | 2 | . |
| NN | 2 | . | GTATTGTTCAACCAGTGAAA | 20 | . | TCATCAAAATGAAGACTACA | 877 | . |
| TGGCTCTAAAACAAAACCCT | 461 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | GGAAGTACCAGTGGTGTGAG | 1191 | . | CAACCAGATGCTGGTTTATT | 930 | . |
| ACCTTAGACTAAAGCGGACA | 507 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | CTCCACCTCTGCTGTCTCAG | 945 | . | AAATTGTTTTGATAAAGCTC | 1119 | . |
| TCCTCCACAAAGAATAATTA | 478 | . | NNNNNNNNNNNNNN | 14 | . | N | 1 | . |
| N | 1 | . | TATCACATCTTGATGAGAGC | 709 | . | AAAGGAAGTGGCATTGAACT | 373 | . |
| AGAAATTCGTCTTGATTTCC | 526 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | CCCATGCTATTTCTTTATCC | 1307 | . | GTCAAAACTACTAGATGGCA | 71 | . |
| TTCCAAAAAAGAAAAACGAA | 652 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | CAACCAGATGCTGGTTTATT | 887 | . | CTTATGTTAGCAAAAGTATG | 122 | . |
| GGTCTGGTAGATAAAGCATG | 995 | . | NNNNNNNNNNNNNNNNNNN | 19 | . | NNNN | 4 | . |
| NN | 2 | . | GTCTACCTATTTGCTAGAGT | 92 | . | AATTTATTAAACAATTTTAT | 242 | . |
| TCATCAAAATGAAGACTACA | 1766 | . | NNNN | 4 | . | NNNN | 4 | . |
| NNNNNNNNNNNNNNNNNNN | 19 | . | CTTAAGTTACATATACTTTG | 325 | . | TTTAAATTATTATGTTCCCA | 95 | . |
| GTCTACTTATTTGCTAGAGT | 24 | . | N | 1 | . | NNNNNNNNNNNNNNNNNNNN | 595 | . |
| N | 1 | . | CCCCTAAGTTATCCTCCATA | 63 | . | AAAAGAGGTTTAATTGCCTC | 1058 | . |
| CAATTGTTTTGATAAAGCTC | 1376 | . | NNNNNNNNNNNNNNNNNNNN | 3084 | . | NNNNNN | 6 | . |
| NNNNNNNNNNNNNNNNNNNN | 54 | . | CTTCATCTATCTAGAAATAA | 119 | . | ATTTGTTGAACTTCAAAATC | 1119 | . |
| GGTAGTCCCTAACATCGTGC | 308 | . | NN | 2 | . | NNNNNNN | 7 | . |
| NNNNNNNNNNNNNNNNNNNN | 144 | . | AATGTCTTTCTGTAATTTTA | 978 | . | GAGACGGAGTCTCGTTCTGT | 349 | . |
| GATTAATTTATAATTAATTA | 34 | . | NNNNN | 5 | . | NN | 2 | . |
| NNNNNN | 6 | . | TAAAATTTTATAAAATAACT | 207 | . | CCACATACACA | 11 | . |
| GTTAAATTTATTAATTACTT | 2891 | . | NN | 2 | . | NNNN | 4 | . |
| NNNNNN | 6 | . | CAGTGGCGTGATCTCGGCTC | 342 | . | TATTTAAACTAATTTCATTC | 239 | . |
| GAGACAGAGTCTCGCTCTGT | 36 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | CGTTCTCACAATGACATTTT | 1143 | . | GGAATTTGTCTTTTGTGATA | 815 | . |
| CAGTGGCGTGATCTCGGCTC | 311 | . | NNN | 3 | . | N | 1 | . |
| NN | 2 | . | TCATACTGCTAATGAAGTTT | 371 | . | GTCTTTCATGATACACTTTA | 102 | . |
| CCACATACACACACATATTT | 254 | . | N | 1 | . | NNN | 3 | . |
| N | 1 | . | TAAGAGAGCTTCATGACATA | 200 | . | TCATAGTGTTAATGAAGTTT | 670 | . |
| GGAATTTGTCTTTTGCAATA | 240 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | AATCAGAAAATTGATAGCCG | 516 | . | ATATTTTTTAAAAATGCATT | 286 | . |
| ATTTTAGGGGACCTGTCACT | 1252 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | ATTGCTTGTTTCAAATGCCT | 479 | . | GTCTTACTTATTTGGACTTA | 315 | . |
| AATCAGAAAATTGATAGTCG | 97 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GGAAATAGAAATACATAATT | 462 | . | AGAATTTTCAGAGGCCAAGC | 889 | . |
| AAATTTTTTAAAAATGCATT | 418 | . | NNNNNNN | 7 | . | NNNN | 4 | . |
| N | 1 | . | GTGAATAATTTAAATATTAA | 565 | . | TTGATTTAAGTTGTTTCATT | 318 | . |
| ATTGCTTGTTTCAAATGCCT | 40 | . | N | 1 | . | NN | 2 | . |
| NNNNN | 5 | . | TGCGACCTTGGACAAAAGAC | 166 | . | TTTTTTCCCCATATCTGAGG | 271 | . |
| GAAGAGCATATGTTGTTAAC | 139 | . | N | 1 | . | NNNNN | 5 | . |
| N | 1 | . | TATTGCTATGTGGTCATTTA | 104 | . | TATTTTT | 7 | . |
| AGAACTTTGGGAGGCTGAGC | 418 | . | NNNNNNNNNNNNNNNNNNNN | 24 | . | N | 1 | . |
| NNNNNNN | 7 | . | GGAACGATCCTCTTAATTTC | 692 | . | CAATTGGATGTGGTCATTTA | 421 | . |
| ATATAAAAAGCAAGGCATTT | 464 | . | NNNN | 4 | . | N | 1 | . |
| NNNN | 4 | . | ATAAAACTTTTGCCTTTCTT | 103 | . | GTCTAACTTGCTTTCACATA | 352 | . |
| TTGATTTCAATTGTTTCATT | 318 | . | N | 1 | . | NNNN | 4 | . |
| NN | 2 | . | TTTTTTAAATTTTTTTTT | 18 | . | ACCCTGTCTCAATAAATATA | 171 | . |
| TTTTTTCCCCATATCTGAGG | 116 | . | NNNNNNNN | 8 | . | NNNNNN | 6 | . |
| N | 1 | . | GAGACAGAG | 9 | . | AGACAGAG | 8 | . |
| TGCGACCTTGGACAAAAGAC | 166 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TTTCCCTCTGTCACCCAGGC | 230 | . | TTTCCCTCTGTCACCCAGGC | 230 | . |
| AGTTGCTATATGGTCATTTA | 104 | . | N | 1 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 24 | . | TTGGCCTCCAAAGTGCTGGG | 26 | . | TTGGCCTCCAAAGTGCTGGG | 26 | . |
| GGAACGATCCTCTTAATTTC | 215 | . | NNNNNNNNNNNNNNNNNNNN | 16036 | . | NNNNNNNNNNNNNNNNNNNN | 16036 | . |
| N | 1 | . | GGTGTGAGCCACTGCGCCCG | 68 | . | GATGTGAGCCACTGCGCCTG | 68 | . |
| GCATTTGCCCAATTAAAGGG | 17100 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | AGGTACCCTTTAGAAGCCTC | 1022 | . | AGGTACCCTTTAGAAGCCTC | 118 | . |
| GTCTTTTTATCACAGGAATG | 161 | . | NNN | 3 | . | N | 1 | . |
| NNNN | 4 | . | TGTCTGAGGTCTCTTTTCTC | 518 | . | GTCTTTTTATCACAGGCATG | 161 | . |
| ATTACTGTATTCTCTACTTC | 1333 | . | NNNNNNNNNN | 10 | . | NNNN | 4 | . |
| N | 1 | . | CATATTGCAACAAGACAGTT | 432 | . | ATTACTGTATTCTCTACTTC | 1313 | . |
| GGGGTTTTACTTATAATTGA | 1605 | . | N | 1 | . | NNNN | 4 | . |
| NN | 2 | . | CAGAATCAGCTGATGAACAT | 1471 | . | GTTACATTAGTCTACATGGG | 1246 | . |
| TCAATTTTTTAAATTTGCTT | 732 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | ACTTCACTGTTCACAGTGTA | 397 | . | TTTCTTAGGAACTTTGTTCT | 1010 | . |
| GTCTTTGTAAGCTTTTCTCC | 825 | . | N | 1 | . | NNNNNNNNN | 9 | . |
| NNNNN | 5 | . | TACTCACGTAGGGATAGGGC | 22 | . | GGATAGGGCTA | 11 | . |
| ATAAATTACACAGTCTTGGG | 43 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TATTGGCAAAATGCCGCCCG | 102 | . | TATTGGCAAAATGCCGCCCG | 2653 | . |
| CAGACTAATACAGTAAGTTG | 89 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | ATTTTCCCTTTCCCGTTCCT | 969 | . | GCAATGCTTCATAAACAAAG | 508 | . |
| CCCGTGAAAGCAGCCGAGAG | 25 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TTGTACTATGCAAAGCCACA | 1280 | . | CAGATAAGTGAAAATAGTAA | 116 | . |
| TTGTACCATGCAAAGCCACA | 1624 | . | NNNNNNNNNNNN | 12 | . | NNNNNNNNNNNNN | 13 | . |
| NNNN | 4 | . | ATCTGAACTTCTTTATTGTA | 294 | . | AGTATAGAATAAACAATAAT | 490 | . |
| CTATGGTGCACATATTGGCC | 578 | . | NNNNNNN | 7 | . | N | 1 | . |
| NNNNNNNNNNNNN | 13 | . | AAACAAAGTTTTATGGCAGA | 31 | . | TTCACTGTATTCAGAAAATT | 302 | . |
| AGTATAGAATAAACAATAAT | 264 | . | NNNN | 4 | . | NNNN | 4 | . |
| NN | 2 | . | CTATGGTGCACATATTGGCC | 197 | . | GTACTGTAACTTTACTAAAA | 270 | . |
| TAATAGTCAGAAAAATTATG | 224 | . | NNN | 3 | . | NNN | 3 | . |
| N | 1 | . | AAACTGCAAGGAGAGATATG | 261 | . | GATAGGTAAAGAGGCTGTAA | 112 | . |
| TTCACTGTATTCAGAAAACT | 302 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | CAGATAAGTGAAAATGGTAA | 826 | . | AAAAAGGAATCATCTTGGAA | 935 | . |
| GTACTGTAACTTTACTAAAA | 283 | . | NN | 2 | . | NNNNNNNNNNNNNNNN | 16 | . |
| NNNNNNNNNNN | 11 | . | GGTAAGTCTACTTTTATTTG | 1419 | . | GTTGGAGCCTGAGGTTTAGG | 505 | . |
| AATACAGAAATTTCATGTTG | 1718 | . | NNNNNNNNNNNNNNNN | 16 | . | NNNN | 4 | . |
| NNNN | 4 | . | CATGGAGCCTGAGGTTTAGG | 45 | . | ATATGAACTCAAGAACTTAG | 46 | . |
| ATTTACCCTCCTGCAGGAAA | 37 | . | N | 1 | . | NNNN | 4 | . |
| N | 1 | . | CGTGGCAGGCCAAAAAGCAA | 256 | . | AAAAAAAGTAGACAAAAGGC | 562 | . |
| CAAAAGATATGAAAAAAATG | 1134 | . | NNNNN | 5 | . | NNNN | 4 | . |
| N | 1 | . | GTTGATGGACTCACCCTTTC | 114 | . | CAAGGCATACAGAGTGATTT | 46 | . |
| ATTTTGAAGGGACTTGTAAA | 120 | . | NNNN | 4 | . | NNN | 3 | . |
| N | 1 | . | GCCTAACATATATAAAGTAA | 133 | . | ACCTTTCTACAACCTAGCCC | 71 | . |
| ACCTTCTGACCATGCTGATA | 65 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | AATAAAAGTAGACAAAAGGC | 116 | . | GCTCCTGACTCAGGTAAAA | 19 | . |
| AAAAAAGAATTGCCTTTACA | 554 | . | NNNN | 4 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNNNNN | 24 | . | ATTTACCCTGCTGCAGGAAA | 566 | . | AAAAGATACTCAGGCTGTTG | 508 | . |
| GCCATTCCCTATTTATTCAT | 188 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GTTCCTGACTCAGGTAAAAG | 529 | . | CAAAGCTGTACATGTAGATA | 196 | . |
| TGGTGTTCATTGCACCTCGC | 280 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | CAAAGCTGTACATGTAGATA | 75 | . | ACCTTCTGACCATGCTGATA | 65 | . |
| TCCAAAGAAGTTCAAAATTA | 43 | . | N | 1 | . | NNNN | 4 | . |
| N | 1 | . | ATTTTGAAGGGACTTATAAA | 806 | . | AAAACAGGATTGCCTTTAAA | 258 | . |
| AGATATCTA | 9 | . | NNNN | 4 | . | N | 1 | . |
| NNNN | 4 | . | GGCTTAGTGTTTCCCTGCTG | 647 | . | GCCAGCTCTCATCAGAGATA | 295 | . |
| CTTGCTATCCCCCAGATTTC | 99 | . | NNNNNN | 6 | . | NNNNNNNNNNNNNNNNNNNN | 24 | . |
| N | 1 | . | CTGTAATGAACACAGTGCTA | 316 | . | ACCATTCCCTATTTATTCAT | 38 | . |
| GCAATCTCACCTCCTGCCTC | 62 | . | N | 1 | . | NNNN | 4 | . |
| NNNN | 4 | . | ACTTCAATCTCCATCTCAGA | 873 | . | GGCTTAGTGTTTCCCTGCTG | 427 | . |
| CCCTGTAATGAACACAGTGC | 78 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | CACCTCAGCAACCCAACTAA | 127 | . | TCCAAAGAAGTTCAAAATTA | 43 | . |
| GTCATGTCACCAAGGTCACA | 631 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GCATATATATTTGCAACTGA | 38 | . | AGATATCTA | 9 | . |
| TACATTGGTATCCATTAATT | 648 | . | NNNNNNNNNN | 10 | . | NNNN | 4 | . |
| NNNNNNNNNN | 10 | . | AGGACACTTCTTATCTGTTA | 98 | . | CTTGCTTTCCCCCAGATTTC | 99 | . |
| AGGACACTTCTTACCTGTTA | 126 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | TTACTTTTGATTTTTTCAAT | 82 | . | GCAATCTCACCTCCTGCCTC | 435 | . |
| ATTCTCT | 7 | . | NNNNNN | 6 | . | N | 1 | . |
| NN | 2 | . | GAAAATGACATATAGCATTC | 757 | . | TTTTCCTTTATTCTGCTTTA | 340 | . |
| TTTATATATTTCAAAGGTAA | 132 | . | NN | 2 | . | N | 1 | . |
| NNNNNNNNNN | 10 | . | TACATTTTGTGGGGAGCCAC | 438 | . | TACATTGGTATCCATTAATT | 609 | . |
| CAATATAATACATATTGTTT | 162 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | GCTAAGTCCGTAAAGGTATG | 109 | . | GCATATATATTTGCAACTGA | 182 | . |
| GTGTGATTGTTCTCAGCAAG | 376 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | AGGCTCTGCACCAGGTAATA | 102 | . | TTTATATATTTCAAAGGTAA | 139 | . |
| TCTTCTTATAATGACATTAT | 676 | . | NNNNNN | 6 | . | NNNN | 4 | . |
| N | 1 | . | TCTGTTGGTAATGGTCAGTC | 244 | . | AGTATAATACATATTGTTTC | 666 | . |
| AGGCTCTGCACCAGGTAATA | 547 | . | NNN | 3 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNNNNN | 25 | . | TCACAAAAAATCTTGGCTTT | 163 | . | TACATTTTGGGGGGAACCAC | 448 | . |
| GATGTGGCTTTAAAATTAAC | 300 | . | NNNNNNNNNNNNNNNNNNNN | 54 | . | N | 1 | . |
| N | 1 | . | GATGTGGCTTTAAAAGTAAC | 52 | . | TAAAGGTATGCTCCCAGAGT | 202 | . |
| TAGGAAGAATTGACTTTTCA | 367 | . | N | 1 | . | NNNNNN | 6 | . |
| N | 1 | . | GTTCCAAACACAAATTTTTC | 897 | . | TCTATTGGTAATGGTTAGTC | 516 | . |
| AAGGCACATCCACCGTGGTG | 1164 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | TTCATATGCCTATTCTATTC | 84 | . | GTTCCAAACACAAATTTTTC | 247 | . |
| GGTTCCACATTATAAACCTA | 117 | . | NN | 2 | . | N | 1 | . |
| NNN | 3 | . | ATAAGCATTTCTCAAGTATT | 796 | . | TAGGAAGAATTGACTTTTCA | 354 | . |
| AACTATGTGCAAAGAGATAC | 349 | . | NNNN | 4 | . | NNN | 3 | . |
| N | 1 | . | GGTTCCACATTATAAACCTA | 117 | . | TGTGTCATGACAAGGCACAT | 349 | . |
| CCTAAGATTATGCCAGCTAG | 397 | . | NNN | 3 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 46 | . | AACTATGTGCAAAGAGATAC | 349 | . | TTGCTTGGAAAAATAGATGA | 1103 | . |
| CCTTCATGCCCTTTAGGACA | 420 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | CCTAATATTATGCCAGCTAG | 381 | . | TAACAGTAA | 9 | . |
| CCTCACTTGCAAGAACAGAT | 1161 | . | NNNNNNNNNNNNNNNNNNNN | 60 | . | NNNNN | 5 | . |
| N | 1 | . | CTCCTTCATGCCCTTTAGGA | 65 | . | AAATAAAAAGGCTATACAAG | 132 | . |
| CCCAGCTTAACTGAGGGAAA | 377 | . | NNNN | 4 | . | NNNNN | 5 | . |
| N | 1 | . | AGGAAATCAGTTCTCTAGAA | 555 | . | ACATCTTCTGCTAACAATTT | 971 | . |
| CCTCCTCTGCCCCTGTTCCC | 86 | . | NNNNNNN | 7 | . | NNN | 3 | . |
| NNNN | 4 | . | GCCCTCGACTGGGAGGAATT | 175 | . | CAAGATGACAACGATATCTC | 137 | . |
| GACAAGCTCATCTAACTGTG | 962 | . | NNNN | 4 | . | NNNNN | 5 | . |
| N | 1 | . | CCTTGAGTTGTTGACTGCCT | 604 | . | GCCCTCATGTGGGAGGAATT | 175 | . |
| GTGATTTTTCTTAGCTCATC | 263 | . | NNNNNNNN | 8 | . | NNNN | 4 | . |
| NN | 2 | . | GAAACTTATTGAGTTCTGAA | 78 | . | TCTTGAGTCGTTGACTGCCT | 660 | . |
| CAACATGATTTGGGTGATAT | 480 | . | NNNN | 4 | . | N | 1 | . |
| N | 1 | . | GTAAAAGAAAGATATTCAAG | 244 | . | ATTAAAAAATGATGTGAGAG | 113 | . |
| GAAAGTATTTACTAAGCAAA | 220 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | CTGCAAAATAATCTCATAAT | 1265 | . | CCCAGCTTAACTGAGGGAAA | 36 | . |
| CAAGAAACTAGGCATAGAAG | 460 | . | N | 1 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNNNNN | 83 | . | GTGATTTTTCTTAGCTCATC | 543 | . | GTATTTAAGGTGTACAGCAA | 125 | . |
| GTTGGAAAACGGTATGGAGA | 453 | . | N | 1 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNN | 17 | . | GTATCAAATAGCTAAAAAAC | 314 | . | CTGCAAAATAATCTCATAAT | 212 | . |
| ACAAAAAACTACATATTGGG | 875 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | GTGATACACTACATAACCAG | 310 | . | CCTCCTCTGCCCGTGTTCTC | 1596 | . |
| CCTATTCTTCCTCTTTAGGT | 897 | . | NNNN | 4 | . | NNN | 3 | . |
| N | 1 | . | GGGCATCCAAATCAGAAAAG | 253 | . | ATCAAATAGCTAAAAAACAT | 239 | . |
| TTGCTTTTTGTTTTTATTTG | 352 | . | NNNNNNNNNNNNNNNNNNNN | 80 | . | NNNN | 4 | . |
| NN | 2 | . | ATCTATGGAAAACGGTATGG | 265 | . | ATACACCACGATCAAGCGGG | 69 | . |
| TTTCACGTAGCAATAGAACA | 347 | . | NNN | 3 | . | NN | 2 | . |
| NNN | 3 | . | GTCTTCTGCAGCAACTTGGA | 188 | . | GTGATACACTACATAACCAG | 106 | . |
| CTCCGGGTTACTCAACTGAC | 973 | . | NNNNNNNNNNNNNNNNN | 17 | . | NN | 2 | . |
| N | 1 | . | ACAAAAAACTACACATTGGG | 888 | . | AACAAACTAGGTATAGAAGG | 272 | . |
| GCATCATAAATCATTTAAAT | 36 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | TTTATGTGAAACTTTGTACC | 1586 | . | CTAAAAGTAGATCTATCCTT | 647 | . |
| GTCTCCTCTTTCTCCCACAC | 863 | . | NNN | 3 | . | N | 1 | . |
| N | 1 | . | CTCCGGTTTACTCAACTGAC | 1125 | . | GTACACAAAGGCAAATAGAG | 243 | . |
| TCATTTGTTTGAAGGCATGG | 767 | . | NNNN | 4 | . | N | 1 | . |
| N | 1 | . | GCCTATTTTAGCTAAAAGTG | 567 | . | TACAAAATGTATTAATAAGC | 591 | . |
| TCGACAATGGGAAAAGACTC | 624 | . | NN | 2 | . | NNNN | 4 | . |
| NNNN | 4 | . | GTCCATTATACCCTCCTTGC | 1183 | . | CATATTCAACCCCTTGATTA | 1025 | . |
| ATTGGTGGAAGAAAGGAAAA | 1007 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | ACATCAACTGGAAATGTAGG | 385 | . | TTGCTTTTTGGTTTTATTTG | 1677 | . |
| GGAGAAACCCCCTCTCTACT | 194 | . | NNNN | 4 | . | N | 1 | . |
| NNNNNNNNNNNNNN | 14 | . | ATTGTTGGAAGAAAGGAAAA | 58 | . | GCATCACAAACCATTTAAAT | 722 | . |
| GAAAAAGAAAAAAGAAAAGA | 737 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | CAAAAAAGAAAGATGATATG | 948 | . | GTCCATTATACCCTCCTT | 18 | . |
| TACTAAAGGTAAAACAGAAC | 337 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GGAGAAACCCCGTCTCTACT | 413 | . | CCTAGGTCCTCATAATCTAC | 158 | . |
| CAACTGTTAGAACAAAGAGA | 333 | . | NNN | 3 | . | NN | 2 | . |
| N | 1 | . | GAAAGCTGGCATTTTCATAA | 331 | . | CATTTGTTTGAAGGCATGGG | 1453 | . |
| TTGCCCTCAAGGTAAAGATA | 130 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GCCATGTGACTTTGCATTTC | 197 | . | CAAAAAAGAAAGATGATATG | 1143 | . |
| AGTACCCAAAAGAGCAGGGG | 167 | . | N | 1 | . | NNNNNNNNNNNN | 12 | . |
| NNNNNNNNNNNNNNNNNNNN | 64 | . | TACTGAAGGTAAAACAGAAC | 436 | . | AAGAAAAAGAAAAAAGAAAA | 207 | . |
| CCTCAGTTTTCCTCCCTCTC | 305 | . | NN | 2 | . | NNN | 3 | . |
| N | 1 | . | GTACCTTACGAATTTTTGGA | 233 | . | GAAAGCTGGCATTTTCATAA | 331 | . |
| AACTTCAGGGGTGGGAGTTG | 1569 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TTGGCCTCAAGGTAAAGATA | 130 | . | GCCATGTGACTTTGCATTTC | 634 | . |
| ACAGTGCAAGCCAGTTAAAT | 387 | . | N | 1 | . | NN | 2 | . |
| NNNNNN | 6 | . | AGTACCCAAAAGAGCAGTGG | 167 | . | GTACCTTATGAATTTTTGGA | 334 | . |
| TCAGAAAGAGCATGCTCCCT | 35 | . | NNNNNNNNNNNNNNNNNNNN | 64 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 6763 | . | CCTCAATTTTCCTCCCTCTC | 305 | . | AAAATTTGGAATATTGTTCT | 1214 | . |
| TGCAGAAGTGCAGTTAATGC | 845 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | AACTTCAGGGGTGGGAGTTG | 469 | . | CAATTTTTTCAATATATATA | 921 | . |
| CAACTATTCTCCTTGCAGCT | 719 | . | N | 1 | . | N | 1 | . |
| NNNNN | 5 | . | AAATCTTAATAATAATAATG | 1674 | . | ACAGTGAAAGCCAGTTAAAT | 1828 | . |
| AGTCAGAAAAAGAAAGTATT | 1886 | . | N | 1 | . | N | 1 | . |
| NNN | 3 | . | CTACTTTACCCCATATGTTT | 212 | . | CCACCCACTACAGACAAGCC | 1729 | . |
| GAGTCCTCCAAACTTTTCCA | 106 | . | NNNNN | 5 | . | NN | 2 | . |
| NNNNNNNNNNNNNNNNNNNN | 24 | . | AGTCTAAACAATGTGCCCTT | 169 | . | GTCTCCACTAAAAATACAAA | 532 | . |
| ATGAAGAAACATCCAAGACT | 58 | . | NN | 2 | . | N | 1 | . |
| NNNNNNNNN | 9 | . | GCCTCAGAAACACACTTTTG | 1113 | . | AAAGAAATGATAAACGTTTG | 167 | . |
| TTGCATGTGGCTGGGGAGGC | 593 | . | N | 1 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 100 | . | CAGATACTTCCCAAAGGAAG | 204 | . | AAAAACAAACCGGAACAAAA | 1210 | . |
| ACGAAGTTTCACTCTTGTTG | 68 | . | NN | 2 | . | NNNNNNNNNNN | 11 | . |
| N | 1 | . | TCCTAAAATTTGTATGGAAC | 1457 | . | GCTGGGCGTGGTGGCAGGCA | 54 | . |
| TCCTGGGTTCAAGTGATTCT | 327 | . | NNNN | 4 | . | NNNNNNN | 7 | . |
| N | 1 | . | CACACACAAAAAACAATGAA | 607 | . | CCTTGAACCTGGGAGGCAGA | 2299 | . |
| TAACAGAAAAAAAAAAAAAA | 344 | . | N | 1 | . | NNNNNNNNNNN | 11 | . |
| NNNN | 4 | . | TTGCAAGTGCTTATTAAGTT | 1282 | . | CAAGCCCTCTGGAAGAGACT | 317 | . |
| AGTGGGTAAGTGACTCAATC | 63 | . | NNNNNNNN | 8 | . | N | 1 | . |
| N | 1 | . | GACAAACAAACGCCATAAAC | 320 | . | TATACCCACACAGGATTTGA | 101 | . |
| ATGTCATAGTCATGATATAA | 1678 | . | NNNNNNNNN | 9 | . | NNN | 3 | . |
| NNNNNNNN | 8 | . | GAGACGGAGTCTGGCTCTGT | 127 | . | TAGAAGCCACAGTTCTTGGG | 359 | . |
| TTAGCCAGCCGTGGTGGAGC | 625 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | CCCCCCACCACGCCAGGCTA | 222 | . | TTTCTATTCAATATTCAACA | 205 | . |
| GAAGAGAGAAATCCAAAGAG | 121 | . | NNNNNNNN | 8 | . | N | 1 | . |
| NNNN | 4 | . | CCTCAATAAAACAACCACTG | 1245 | . | TAATTTATAAAGAAAAGAAG | 387 | . |
| CTATTATTTTTCCCCTTTGA | 456 | . | NNN | 3 | . | N | 1 | . |
| N | 1 | . | CTCTCATTTGCACCTCTTCT | 265 | . | AAAGACTTAACTCATTCTAG | 252 | . |
| CATGTGCTTCTCCTGATTGG | 481 | . | NNNNNNNNNNN | 11 | . | N | 1 | . |
| N | 1 | . | CAGGGCCTCTGGAAGAGACT | 317 | . | TATTTTGATTCCATGTCCCA | 402 | . |
| TCCTTCTCCTCTCTCTGCCT | 569 | . | N | 1 | . | NNNN | 4 | . |
| N | 1 | . | TATACCCACACAGGATTTGA | 463 | . | GTTCTTGACTTCTGTGCACC | 99 | . |
| TGGGTGGCACCAACCCTTCT | 1009 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TTTCTATTCAATATTCAACA | 3305 | . | CGCTTTTAGCCATGGCTGGA | 850 | . |
| GGGAGAATACCTTGGGAAGA | 278 | . | NNNNNNN | 7 | . | NNNNNNNNN | 9 | . |
| N | 1 | . | GGAAAAAAGAAAAAGCAGAA | 401 | . | TTGCATGTGGCTGGGGAGGC | 593 | . |
| TGGGTGAGGGAAATTTCTGA | 94 | . | N | 1 | . | NNNNNNNNNNNNNNNNNNNN | 100 | . |
| N | 1 | . | ATGTCATAGTCATGACATAA | 51 | . | ACGAAGTTTCACTCTTATTG | 400 | . |
| CAACATACTTATAGCTACTT | 224 | . | NN | 2 | . | NNNNNNNNNN | 10 | . |
| NNNNNNNNNNNNNNNNNNNN | 5253 | . | ACAGTTTCCTCCCTTTCTGA | 299 | . | GGGAAAAAAAAGCAGCACTA | 450 | . |
| TGCCTGGCAGGTCTCTCCCT | 229 | . | NNNN | 4 | . | NN | 2 | . |
| NNN | 3 | . | TTAGCTATTTGATCATTAAG | 27 | . | ACAGTTTCCTCCCTTTCTGA | 256 | . |
| AGAAAAGAAAAGAAAAAGAA | 211 | . | NN | 2 | . | NNN | 3 | . |
| N | 1 | . | GTTTTGTTTTTGTTTTTGTT | 137 | . | GTTTCTTCCCACCTCAGCAA | 59 | . |
| ATGGGACTAATGCCATGTAT | 79 | . | NNN | 3 | . | N | 1 | . |
| N | 1 | . | ATTACAGGTGCACGCCACTA | 29 | . | GTAAGGTTTTT | 11 | . |
| GAGACTAGAACAGAGGTAAT | 184 | . | N | 1 | . | NN | 2 | . |
| NN | 2 | . | TTTTGTATTTTTAGTAGAGA | 1269 | . | GTTTTGTTTTTGTTTTTGTT | 169 | . |
| ACATAATACTTTTAAAAGGA | 443 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GCCCC | 5 | . | TTTTGTATTTTTAGTAGAGA | 251 | . |
| GTGAAATCATTTTAAATCCT | 2418 | . | NNNNNNNNNNNNNNNNNNNN | 98 | . | NN | 2 | . |
| N | 1 | . | TGGAAAAGAAGAAAGAAAGT | 227 | . | CTTAGGTTGCAGTTCCTGCT | 1016 | . |
| TGATATTTTAAAAATTATAT | 309 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | CAAAAAGGAAATGGTAAAGT | 692 | . | GCCCC | 5 | . |
| AGCTTGGATAGCCGATGCCT | 592 | . | N | 1 | . | NNNNNNNNNNNNNNNNNNNN | 154 | . |
| N | 1 | . | GTCTTTGAGAAAGAAAGAAA | 46 | . | AGAAAAAGCAGAAAAAAAGA | 171 | . |
| TATTTGAAATTGCAAAGAGC | 29 | . | N | 1 | . | NN | 2 | . |
| NN | 2 | . | CCTGTGCTTCTCCTGATTGG | 605 | . | AAAAAAGGAAATGGTAAAGT | 279 | . |
| CCCAGACAATATTGACAAAG | 289 | . | NNNNN | 5 | . | NNNN | 4 | . |
| NNNN | 4 | . | GGATAACATTTAAATAGAAT | 492 | . | CTATTATTTTTCCCCTTTGA | 409 | . |
| AGCTGATCACTTGAGGCCAG | 272 | . | NNNNNNNNNNNNNNN | 15 | . | N | 1 | . |
| NNNNNNN | 7 | . | CCCCTGAATTGTATCACCAA | 85 | . | GTCTTTGAGAAAGAAAGAAA | 342 | . |
| TTATGCTGAGTCAGTAGAAT | 262 | . | N | 1 | . | NNNNNNNNNNNN | 12 | . |
| NNN | 3 | . | AACATTTATTGAGCAATTAC | 1232 | . | TAGCACTTTCTTAGCAACAC | 298 | . |
| CAAGAAAGTGAAAGGCAAGT | 873 | . | N | 1 | . | NNNNN | 5 | . |
| NNNNNNNN | 8 | . | CAACATACTTATAGCTGCTT | 224 | . | GGATAACATTTAAATAGAAT | 441 | . |
| CTTTCATGCTCCTTAGGAGA | 174 | . | NNNNNNNNNNNNNNNNNNNN | 5253 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 3399 | . | TGCCTGGCAGGTCTCTCCCT | 523 | . | TGGGTGGCACCAACCCTTCT | 50 | . |
| TGGTGGCGGGTGCCTGTAGT | 164 | . | N | 1 | . | NNNNNNNNNNNNNNN | 15 | . |
| NNNNNNNNNNNNNNNNNNNN | 21 | . | GAGACTAGAACAGAGGTAAT | 184 | . | CCCCTGAATTATATCACCAA | 624 | . |
| TTGATGATCTGATGATGATC | 864 | . | N | 1 | . | NNNN | 4 | . |
| NN | 2 | . | CACATAATACTTTTAAAAGG | 261 | . | ATTTACTTGAAGTGTTCTAT | 518 | . |
| AAAAACACGAAGACAAAAAT | 882 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | TAATTTCAAGTGAAAGCACA | 57 | . | AATCCTGGCAAACTAATACA | 110 | . |
| GTTTCCTTCTGTAAAATAAG | 199 | . | N | 1 | . | NNNNNNN | 7 | . |
| NNNNNNNNNNNNNNNNNNNN | 321 | . | AAAAAAACTAAATGAATATT | 124 | . | ACCACATAAACATTTTTTTC | 5757 | . |
| TTTAAGGACACAGCTATATT | 343 | . | N | 1 | . | NNNNNNNNNNNN | 12 | . |
| NNNNNNNN | 8 | . | GTGAAATCATTTTAAATCCT | 1361 | . | GAAAAAGAAAAAAGAAAAGA | 733 | . |
| CAACTTATAAACATTTTCTT | 148 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | GAGTATTTACAAGATAATCT | 78 | . | TAATTTCAAGTGAAAGCACA | 290 | . |
| GACAATTATTATATTCACGT | 47 | . | NNNN | 4 | . | N | 1 | . |
| NNNNNNNNNNN | 11 | . | TTTCTCAGATGAAATTATCA | 893 | . | AAAAAAATTAGCAGTAAAAG | 188 | . |
| TTCTTTAACTCAAAACAAAC | 793 | . | NN | 2 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 1359 | . | ATTAAATGAATTTAGAAAAA | 568 | . | CTATCAAATATTGAAAAATT | 106 | . |
| AAAATACCATGCCAAACGGT | 517 | . | N | 1 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNN | 17 | . | TAAATACTTGAAGAGCCCCT | 472 | . | TATTACATATCAAAACTTAT | 957 | . |
| TTTTTTATTGAAGTTCTAGG | 143 | . | NN | 2 | . | N | 1 | . |
| N | 1 | . | AGTCAGAGGACCTATACTAC | 257 | . | GAGTATTTACAAGATAATCT | 275 | . |
| ACCCCACAACAGACCCCGGT | 94 | . | N | 1 | . | NNN | 3 | . |
| NN | 2 | . | AA | 2 | . | AAATCTCTAAGTTTCCAATA | 1089 | . |
| GGTTTTTTGTCTTTGCTATA | 460 | . | NNNN | 4 | . | N | 1 | . |
| NN | 2 | . | AGCTGATCACTTGAGGCCAG | 116 | . | GCTTGGATAGCCCATGCCTG | 177 | . |
| CATTGTGGTTTTGATTTGCA | 136 | . | N | 1 | . | N | 1 | . |
| NNNN | 4 | . | TATGTTGGGTGGCGGAAGCA | 125 | . | TAAATACTTGAAGAGCCCCT | 91 | . |
| TTTTCTTGTAAATTTGTTTG | 539 | . | NN | 2 | . | NNNNNNNNNNN | 11 | . |
| NNNNNNNNNN | 10 | . | GAAAAAGAAAAAGAAAAGAA | 28 | . | ACACCTAAAGCCCCTGCCCA | 878 | . |
| TAGACTTGTATAGTTTGAAG | 108 | . | NNNNNNN | 7 | . | N | 1 | . |
| NNNNNN | 6 | . | TTATGCTGAGTCAGCAGACT | 262 | . | AAAAAAAAGAAAA | 13 | . |
| GTTTTTTAGAATGCTGTGAA | 1456 | . | NNN | 3 | . | N | 1 | . |
| NNNNN | 5 | . | CAAGAAAGTGAAAGGCAAGT | 104 | . | AGAAAAGAAAAGAAAAAAGA | 148 | . |
| GTATTAGCCCTTTGGGAATC | 2442 | . | NNN | 3 | . | NNNN | 4 | . |
| NN | 2 | . | TCAAAAAGAAAACTTGATTT | 240 | . | TATTTTTATGACTTTTGGGT | 1050 | . |
| GTCAATTAAACCTCTTTTCT | 894 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GCCGTATGTATGCATATCAT | 525 | . | AGGCACGGTGGCTCACACCT | 4512 | . |
| CTTTCCATTATTTAATTGCT | 28 | . | NNNNNNNN | 8 | . | N | 1 | . |
| NNNNN | 5 | . | CTTTCATGCTCCTTAGGAGA | 174 | . | AAGTTTGAGTGAGAAATCTA | 74 | . |
| GGCTGTTGTGAGTTACTGGA | 973 | . | NNNNNNNNNNNNNNNNNNNN | 3399 | . | NN | 2 | . |
| NNNNN | 5 | . | TGGTGGCGGGTGCCTGTAGT | 165 | . | AAAAACACGAAGACAAAAAT | 60 | . |
| AAAAGATTAATGTGGAAAAT | 32 | . | NNNNNNNNNNNNNNNNNNNN | 20 | . | NNNNN | 5 | . |
| N | 1 | . | TTGATGATCTGATGATGATC | 741 | . | ATTCATTTTTAAATTAAACA | 340 | . |
| CCAGAATTGACAGATTTCTG | 281 | . | N | 1 | . | NNNNNN | 6 | . |
| N | 1 | . | GGAAAGACGCTATGCGTAAA | 47 | . | AAATGGAGTAAAGTTGAGAG | 64 | . |
| GCAGGCATCAGGATGCAAAA | 356 | . | N | 1 | . | NN | 2 | . |
| N | 1 | . | AAGTTTGAGTGAAAAATCTA | 1076 | . | TTTGACTTCTCCCTGTCGTC | 606 | . |
| AAATGAAATTTTAGCATAGA | 1409 | . | NNNN | 4 | . | NNNNNNNNNNNNNNNNNNNN | 321 | . |
| NNNNNNNNNNNNNNNNNNNN | 33 | . | AGCTCTTATTATAATTAATA | 761 | . | TTTAAGGACACAGCTATATT | 980 | . |
| TAAAAATCCTTCTCATCTCT | 299 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | CATTTTCTTCAAATTTTCAT | 79 | . | ACCCCTTTGGCTTCCTTTTC | 220 | . |
| ACTGTGTGACCTGGAAATAT | 214 | . | NNNN | 4 | . | N | 1 | . |
| NNNNNNNNNNNNNNNNNNNN | 33 | . | CATTTAATAAAATGGCAATT | 38 | . | TTTTTGAATTTTGTGGTTGA | 2028 | . |
| CTACCATAAACATGGCTTAA | 103 | . | NNNNN | 5 | . | NNNNNNNNNNNNNNN | 15 | . |
| N | 1 | . | ATTTACATATTAATTGACAA | 62 | . | TTATTATACTGAAAGTACTA | 826 | . |
| CTTTTTGCTTTTATTTTTAA | 712 | . | NNNNNNNNNNNNN | 13 | . | N | 1 | . |
| NNNNNNN | 7 | . | CTCTAACTCAACACAAACAA | 791 | . | GATGGGGCTGTTTGTTTTTT | 686 | . |
| TACTCTGGCCAGGGAGACAG | 57 | . | NNNNNNNNNNNNNNNNNNNN | 1359 | . | N | 1 | . |
| NN | 2 | . | AAAATACCATTCCAAAAGGT | 861 | . | CAAATTCTGTGAAGAAAGTC | 589 | . |
| CAAGGCCGGGGGGGATGTCT | 478 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | ATGGCTGCATAGTATTCCAT | 842 | . | AATGCTTGTTTTATAGAGTC | 1532 | . |
| TACAAATAAAAACTATACAC | 815 | . | NNNNNNN | 7 | . | NN | 2 | . |
| N | 1 | . | AAGGGATCCAGTTTCAGCTT | 88 | . | GTTTGGAACTTCCTAGAGAC | 266 | . |
| CTCAGCTCTCTTCTCTTCAA | 639 | . | N | 1 | . | N | 1 | . |
| NN | 2 | . | CTTGTCAGGTTTGTCAAAGA | 1114 | . | CCTGGCTGCTTTTAAAAACT | 1295 | . |
| CACCTTTATTAAAGGAAATA | 256 | . | NNNNN | 5 | . | NNNNNNNNNNNNNN | 14 | . |
| NN | 2 | . | TTTTGCCGATTCAAAGTTTC | 1250 | . | CACATGCTGTTCTCATGATG | 463 | . |
| TGAGAATGTTTCAAGTCCAG | 153 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | GTTTGGAACTTCCTAGAGAC | 503 | . | ATTTCCAAATAGGCCATTTT | 1067 | . |
| GGGAAGACTGAGTCTGTGTC | 50 | . | N | 1 | . | NNNN | 4 | . |
| NNNNNNN | 7 | . | CTCCCATCACAGGCCTGGAG | 1058 | . | TTTTTTAACACAATTAAAGT | 564 | . |
| GAGAGAGAGAGAGAGACTCC | 102 | . | NNNNNNNNNNNNNN | 14 | . | NNNNN | 5 | . |
| N | 1 | . | CACATGCGGTTCTCATGATG | 1121 | . | AAAAGAGTAATGTGGAAAAT | 32 | . |
| ACTCTTACTCAAACCACTGC | 349 | . | NNN | 3 | . | N | 1 | . |
| NNNNNNN | 7 | . | TTTGCTGTTGTGAGTTACTG | 212 | . | CCAGAATTGACAGCTTTCTG | 1650 | . |
| TATACCTAGGTAACAAACCT | 1194 | . | N | 1 | . | N | 1 | . |
| N | 1 | . | TTGAGGATCTTGTGCTAAGT | 129 | . | CTAATACAGTCAATTCCTTG | 729 | . |
| CATTTATGACTTGCCTATTC | 906 | . | N | 1 | . | N | 1 | . |
|  |  |  | TGGGAGAAGGAAGAAATGGG | 64 | . | ACTGTGTGACCTGGAAATAT | 214 | . |
|  |  |  | NN | 2 | . | NNNNNNNNNNNNNNNNNNNN | 33 | . |
|  |  |  | GTTCTTTACCACAATTAAAA | 885 | . | CTACTATAAACATGGCTTAA | 888 | . |
|  |  |  | N | 1 | . | N | 1 | . |
|  |  |  | GCAGGCATCAGGATGCAAAA | 1129 | . | GGGGAGGATGTCTACTGTAT | 75 | . |
|  |  |  | N | 1 | . | NNNNNNNNNNNNNNNNNNNN | 32 | . |
|  |  |  | TGACTGTGAAGATGGAGAAA | 121 | . | CACATATTTATTTCTTACAG | 1941 | . |
|  |  |  | NN | 2 | . | N | 1 | . |
|  |  |  | ATTATTTTTGCCCAGTGAGA | 513 | . | CCACTCCTGATGCAAAAGAG | 343 | . |
|  |  |  | NNNNNNNNNNNNNNNNNNNN | 33 | . | NNN | 3 | . |
|  |  |  | TAAAAATCCTTCTCATCTCT | 642 | . | AAAAAAGAGAGAGAGAGAGA | 106 | . |
|  |  |  | N | 1 | . | N | 1 | . |
|  |  |  | CGTTTTTTCTTTTTGCTTTT | 187 | . | ACTCTTACTCAAACCACTGC | 247 | . |
|  |  |  | N | 1 | . | N | 1 | . |
|  |  |  | CAAAAGTGGGCACTTTGCAG | 532 | . | GGGGCCTGTCGGGGATGGGG | 101 | . |
|  |  |  | NNNNNNNNNNNNNNNNNNNN | 35 | . | NNNNNNN | 7 | . |
|  |  |  | CACAAAAAAAAAAAAAAAAA | 37 | . | TATACCTATGTAACAAACCT | 69 | . |
|  |  |  | N | 1 | . | N | 1 | . |
|  |  |  | GGGGGGGATGTCTACTATAT | 75 | . | GAAACTCATCTCAGCTTCTG | 1013 | . |
|  |  |  | NNNNNNNNNNNNNNNNNNNN | 34 | . | N | 1 | . |
|  |  |  | CATATTTATTTCTTACAGTT | 362 | . | GGCCCCTGGCTGATTTTCTT | 1017 | . |
|  |  |  | NNNN | 4 | . |  |  |  |
|  |  |  | AAAATAAAAACTACACACAA | 396 | . |  |  |  |
|  |  |  | N | 1 | . |  |  |  |
|  |  |  | GTCACAGAAAGACAAATTCT | 606 | . |  |  |  |
|  |  |  | NNNN | 4 | . |  |  |  |
|  |  |  | CCCACACAGAAAGAGAGAGA | 272 | . |  |  |  |
|  |  |  | NNNN | 4 | . |  |  |  |
|  |  |  | CTTCCTATTAGTCTTTCCCT | 170 | . |  |  |  |
|  |  |  | N | 1 | . |  |  |  |
|  |  |  | TCACCTTTCTTAAAGGAAAA | 70 | . |  |  |  |
|  |  |  | NNNNNNNNNN | 10 | . |  |  |  |
|  |  |  | TTAAGCGGAAGTTTGGGGTT | 524 | . |  |  |  |
|  |  |  | NNNNNNNNNNN | 11 | . |  |  |  |
|  |  |  | TGTCCTGCTGATCTCTCATA | 205 | . |  |  |  |
|  |  |  | N | 1 | . |  |  |  |
|  |  |  | GGGGCCTGTCGAGGGTGGGG | 177 | . |  |  |  |
|  |  |  | N | 1 | . |  |  |  |
|  |  |  | GAAACTCATCTCAGCTTCTG | 1013 | . |  |  |  |
|  |  |  | N | 1 | . |  |  |  |
|  |  |  | GGCCCTTGGCTGATTTTCTT | 1017 | . |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**References supplementary material**

s1 The International Genome Sample Resource (IGSR). Providing ongoing support for the 1000 Genomes Project data. <https://www.internationalgenome.org/>. Accessed 13 November 2019.

s2 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

s3 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.

s4 Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12(4):656-64.

s5 Spruyt M, Buquicchio F. Gene Runner http://www.generunner.net/ Accessed 05 July 2019.

s6 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7.

s7 Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

s8 Meyer S, Vollmert C, Trost N, Sigurdardottir S, Portmann C, Gottschalk J, et al. MNSs genotyping by MALDI-TOF MS shows high concordance with serology, allows gene copy number testing and reveals new St(a) alleles. Br J Haematol. 2016;174(4):624-36.

s9 Meyer S, Trost N, Frey BM, Gassner C. Parallel donor genotyping for 46 selected blood group and 4 human platelet antigens using high-throughput MALDI-TOF mass spectrometry. Methods Mol Biol. 2015;1310:51-70. doi: 10.1007/978-1-4939-2690-9\_5.

s10 Harteveld CL. State of the art and new developments in molecular diagnostics for hemoglobinopathies in multiethnic societies. Int J Lab Hematol. 2014;36(1):1-12.

s11 Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 1989;17(7):2503-16.

s12 Wu DY, Ugozzoli L, Pal BK, Wallace RB.Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia.Proc Natl Acad Sci U S A. 1989;86(8):2757-60.