Hepatitis B Virus (HBV) amino acid alignments and genotype-specific consensus sequences

Philippa C. Matthews
Wellcome Trust Clinical Research Fellow, Nuffield Department of Medicine, University of Oxford,
Peter Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY,UK
Honorary Consultant in Microbiology and Infectious Diseases, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
Email: p.matthews@doctors.org.uk

Background:

Hepatitis B Virus (HBV) is the prototype human hepadnavirus. Unified reference sequences are important for informing phylogenetic analysis, studying diversity within and between hosts, identifying sequence changes (sites of polymorphisms / insertions / deletions) and providing standardised sequence numbering. We have developed a database of HLA Class I epitopes within HBV, 'hepitopes', (on-line at http://www.expmedndm.ox.ac.uk/hepitopes), for which we required a unified approach to HBV sequence numbering in order to provide a reference by which epitopes can be located. This work provides aligned and numbered HBV sequences, subdivided by genotype where possible, in the form of a pdf file and an xls spreadsheet.

Methods:

1. Recently published HBV reference sequences[1] were used as a baseline. The sequences were retreived from Genbank (https://www.ncbi.nlm.nih.gov/) where they can be found under the following accession numbers:
   - Pol: AJW31599
   - Large HBs: AJW31600
   - HBx:AJW31597
   - Pre-core: AJW31591 (first 29 residues only, as remainder of sequence is Core protein)
   - Core: AJW31598
2. Amino acid sequences were downloaded from The Hepatitis B Virus Database (https://hbvdb.ibcp.fr) for large HBs, HBx, Core and Pol, for all proteins for which >100 sequences were available. (Download dates 30-Sept to 18-Oct 2016; see Table for numbers of sequences).
3. Downloaded sequences were converted to fasta format using https://hcv.lanl.gov/cgi-bin/FORMAT_CONVERSION/convert.cgi
4. Sequences were aligned using clustal omega (up to a maximum of 2000 sequences per alignment) at http://www.ebi.ac.uk/Tools/msa/clustalo/
5. A consensus sequence and entropy data was generated from each alignment using tools available on-line at the Los Alamos website https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html
6. Aligments for sequences for Genotype E Pol and the first 29 residues of genotype D pre-core protein failed to align, so these are not represented in the consensus sequences.

Table: Number of sequences informing consensus for each HBV protein / genotype

Protein	Geno A	Geno B	Geno C	Geno D	Geno E	Geno F
Pol	1020	1403	1869	982	272	229
Large surface	1271	1799	1897	1167	292	237
HBx	862	1224	2000	1058	288	340
Pre-core	984	1149	1738	768	313	219
Core	1371	1500	1991	1433	453	252

References:

1. Liu WC, Lin CP, Cheng CP, Ho CH, Lan KL, Cheng JH, et al. Aligning to the sample-specific reference sequence to optimize the accuracy of next-generation sequencing analysis for hepatitis B virus. Hepatol Int. 2016 Jan;10(1):147-57.

HBV Pol Amino Acid Alignment (page 1/5)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Terminal protein | Spacer | Reverse transcriptase | RNAse H

Position	1 2 3 4 5 6 7 8 9 10	11 12 13 14 15 16 17 18 19 20	21 22 23 24 25 26 27 28 29 30	31 32 33 34 35 36 37 38 39 40	41 42 43 44 45 46 47 48 49 50
Reference	M P L S Y Q H F R K	L L L L D E E A G P	L E E E L P R L A D	E G L N R R V A E D	L N L G N P N V S I
Geno A	- - - - - - - - - -	- - - - - D - - - -	- - - - - - - - - -	A D - - - - - - - -	- - - - - L - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - L - - - -
Geno C	- - - - - - - - - -	- - - - - D - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - L - - - -
Geno D	- - - - - - - - - R	- - - - - D - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - L - - - -
Geno F	- - - - - P - - - -	- - - - - D - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - Q L - - - - -

Position	51 52 53 54 55 56 57 58 59 60	61 62 63 64 65 66 67 68 69 70	71 72 73 74 75 76 77 78 79 80	81 82 83 84 85 86 87 88 89 90	91 92 93 94 95 96 97 98 99 100
Reference	P W T H K V G N F T	G L Y S S T V P C F	N P K W Q T P S F P	D I H L Q E D I V D	R C E Q F V G P L T
Geno A	- - - - - - - - - -	- - - - - - - - I -	- - E - - - - - - -	K - - - - - - - I N	- - Q - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - K - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - - - V -	- - E - - - - - - -	H - - - - - - - I N	- - Q - Y - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - V -	- - H - K - - - - -	N - - - H Q - - I K	K - - - - - - - - -
Geno F	- - - - - - - - - -	- - - - - - - - A -	- - D - L - - - - -	- - H Q - L I S K	- - - - - - - - - -

Position	101 102 103 104 105 106 107 108 109 110	111 112 113 114 115 116 117 118 119 120	121 122 123 124 125 126 127 128 129 130	131 132 133 134 135 136 137 138 139 140	141 142 143 144 145 146 147 148 149 150
Reference	V N E T R R L K L I	M P A R F Y P N V T	K Y L P L D K G I K	P Y Y P E H V V N H	Y F Q T R H Y L H T
Geno A	- - - K - - - - - -	- - - - - - - T H -	- - - - - - - - - -	- - - - D Q - - - -	- - - - - - - - - -
Geno B	- - - N - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - K - - - - - -	- - - - - - - - L -	- - - - - - - - - -	- - - - - - A - - -	- - K - - - - - - -
Geno D	- - - K - - - Q - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - L - - -	- - - - - - - - - -
Geno F	K - - L - - - - - V	- - - - - F - K - -	- - F - M E - - - -	- - - - - - - - - -	- - K - - - - - - -

Position	151 152 153 154 155 156 157 158 159 160	161 162 163 164 165 166 167 168 169 170	171 172 173 174 175 176 177 178 179 180	181 182 183 184 185 186 187 188 189 190	191 192 193 194 195 196 197 198 199 200
Reference	L W K A G I L Y K R	E S T R S A S F C G	S P Y S W E Q D L Q	H G R L V F Q T S K	R H G D K S F C P Q
Geno A	- - - - - - - - - -	- T - - - - - - - -	- - - - - - - E - -	- - - - I K - - Q -	- - - - E - - - S -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- T - - - - - - - -	- - - - - - - E - -	- - - - - Q - - T -	- - - - E - - - S -
Geno D	- - - - - - - - - -	- T - H - - - - - -	- - - - - - - E - -	- -	. . . A E - - H Q -
Geno F	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - E - -	- - S T S L N D - -	G - - T E - L - A -

➔ *continues on next page*

HBV Pol Amino Acid Alignment (page 2/5)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Terminal protein | Spacer | Reverse transcriptase | RNAse H

Position	201 202 203 204 205 206 207 208 209 210	211 212 213 214 215 216 217 218 219 220	221 222 223 224 225 226 227 228 229 230	231 232 233 234 235 236 237 238 239 240	241 242 243 244 245 246 247 248 249 250
Reference	S P G I L R G S S V	G P S I Q S Q L R K	S R L G P Q P A Q G	Q L A G R Q Q G G S	G S I R A R V H P S
Geno A	P S - - - S R - - -	- - C - R - - - K Q	- - - - L - - H - -	P - - S S Q P - R -	- - - - - - - A - - -
Geno B	- - - - - P R - - -	- - C - - - - - - -	- - - - - - - - - -	- - - - R - - - - -	- - - - - - - - - -
Geno C	- S - - - S R - P -	- - C V R - - - K Q	- - - - L - - Q - -	S - - R G K S - R -	- - - - - - - - - T
Geno D	- S - - - S R P P -	- S - L - - K H - -	- - - - L - S Q - -	H - - R - - - - R -	W - - - - G I - - T
Geno F	- S - - - S R P - A	- S - - - G K F Q Q	- - - - L - Q K - -	- - - N G K - - R -	- R L - S - - - T P

Position	251 252 253 254 255 256 257 258 259 260	261 262 263 264 265 266 267 268 269 270	271 272 273 274 275 276 277 278 279 280	281 282 283 284 285 286 287 288 289 290	291 292 293 294 295 296 297 298 299 300
Reference	P W G T V G V E P S	G S G Q T H N C A S	S S S S C L H Q S A	V R K A A Y S L I S	T S K G H S S S G H
Geno A	T R R Y F - - - - -	- - - H I D H S V N	N - - - - - - - - -	- - - - - - - H L -	- - - R Q - - - - -
Geno B	- - - - - - - - - -	- - - P T - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	T R R S F - - - - -	- - - H I D - S - -	- T - - - - - - - -	- - - T - - - H L -	- - - R Q - - - - -
Geno D	A R R P F - - - - -	- - - H T T - L - -	K - A - - - Y - - P	- - - - - - P A V -	- F E K - - - - - -
Geno F	T R W P - - - - - -	- T - C A N - L - -	R - A - - F - - - -	- - E K - N P S L -	- - - R - T - T - -

Position	301 302 303 304 305 306 307 308 309 310	311 312 313 314 315 316 317 318 319 320	321 322 323 324 325 326 327 328 329 330	331 332 333 334 335 336 337 338 339 340	341 342 343 344 345 346 347 348 349 350
Reference	A V E L H H F P P N	S S R S Q S P G P V	P S C W W L Q F R N	S E P C S E Y C L C	H I V N L I D D W G
Geno A	- - - F - C L - - S	- A G - - - Q - S -	F - - - - - - - - -	- K - - - - - - - S	- L - - - R - - - -
Geno B	- - - - - - - - - -	- - - - - - Q - - -	L - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - N I - - S	- A - - - - E - - I	L - - - - - - - - -	- K - - - D - - - T	- - - - - L - - - -
Geno D	- - - - - N L - - -	- A - - - - E R - -	F P - - - - - - - -	- K - - - D - - - S	- - - - - L - - - -
Geno F	- - - - N S V - - G	- V - - E G K - S -	F - - - - - - - - D	T - - - - D - - - S	- - I - - L - - - -

Position	351 352 353 354 355 356 357 358 359 360	361 362 363 364 365 366 367 368 369 370	371 372 373 374 375 376 377 378 379 380	381 382 383 384 385 386 387 388 389 390	391 392 393 394 395 396 397 398 399 400
Reference	P C T E H G E H L I	R T P R T P A R V T	G G V F L V D K N P	H N T T E S R L V V	D F S Q F S R G N T
Geno A	- - D - - - - - H -	- I - - - - - - - -	- - - - - - - - - -	- - - A - - - - - -	- - - - - - - - - I -
Geno B	- - - - - - - - R -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - N -	- I - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - S -
Geno D	- - A - - - - - H -	- I - - - - - - - -	- - - - - - - - - -	- - - A - - - - - -	- - - - - - - - - Y
Geno F	- - Y - - - Q - H -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - T -

➔ *continues on next page*

HBV Pol Amino Acid Alignment (page 3/5)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation. YMDD motif is marked with a box (residues 549-552 in this alignment).

Terminal protein | Spacer | Reverse transcriptase | RNAse H

Position	401 402 403 404 405 406 407 408 409 410	411 412 413 414 415 416 417 418 419 420	421 422 423 424 425 426 427 428 429 430	431 432 433 434 435 436 437 438 439 440	441 442 443 444 445 446 447 448 449 450
Reference	R V S W P K F A V P	N L Q S L T N L L S	S N L S W L S L D V	S A A F Y H L P L H	P A A M P H L L V G
Geno A	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - I - - -	- - - - - - - - - I -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	H - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - I - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno F	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -

Position	451 452 453 454 455 456 457 458 459 460	461 462 463 464 465 466 467 468 469 470	471 472 473 474 475 476 477 478 479 480	481 482 483 484 485 486 487 488 489 490	491 492 493 494 495 496 497 498 499 500
Reference	S S G L S R Y V A R	L S S N S R I I N N	Q H R T M Q N L H D	S C S R N L Y V S L	M L L Y Q T Y G R K
Geno A	- - - - - - - - - -	- - - - - - - N - -	- Y G - - - - - - -	- - - - Q - - - - -	- - - - K - - - W -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - N	- - - - - - - - - -	- - - - K - - - - -
Geno C	- - - - P - - - - -	- - - T - - N - - Y	- - G - - - D - - -	- - - - - - - - - -	L - - - K - F - - -
Geno D	- - - - - - - - - -	- - - - - - - F - H	- - G - - - - - - -	- - - - - - - - - -	L - - - - - F - - -
Geno F	- - - - - - - - - -	- - - T - - - H D H	- - G - - - - - - N	- - - - - - - - - -	L - - F - - L - - -

Position	501 502 503 504 505 506 507 508 509 510	511 512 513 514 515 516 517 518 519 520	521 522 523 524 525 526 527 528 529 530	531 532 533 534 535 536 537 538 539 540	541 542 543 544 545 546 547 548 549 550
Reference	L H L Y S H P I I L	G F R K I P M G V G	L S P F L L A Q F T	S A I C S V V R R A	F P H C L A F S Y M
Geno A	- - - - - - - - V -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno F	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -

Position	551 552 553 554 555 556 557 558 559 560	561 562 563 564 565 566 567 568 569 570	571 572 573 574 575 576 577 578 579 580	581 582 583 584 585 586 587 588 589 590	591 592 593 594 595 596 597 598 599 600
Reference	D D V V L G A K S V	Q H L E S L Y A A V	T N F L L S L G I H	L N P H K T K R W G	Y S L N F M G Y V I
Geno A	- - - - - - - - - -	- - R - - - - T - -	- - - - - - L - - -	- - - N - - - - - -	- - - - - - - - I -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - L - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - F T S I	- - - - - - L - - -	- - - N - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - F T - -	- - - - - - L - - -	- - - N - - - - - -	- - - H - - - - - -
Geno F	- - L - - - - - - -	- - - - - - - T - -	- - - - - - - - - -	- - T S - - - - - -	- T - H - - - - - -

➔ *continues on next page*

HBV Pol Amino Acid Alignment (page 4/5)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Terminal protein | Spacer | Reverse transcriptase | RNAse H

Position	601 602 603 604 605 606 607 608 609 610	611 612 613 614 615 616 617 618 619 620	621 622 623 624 625 626 627 628 629 630	631 632 633 634 635 636 637 638 639 640	641 642 643 644 645 646 647 648 649 650
Reference	G S W G T L P Q E H	I V Q K I K M C F R	K L P V N R P I D W	K V C Q R I V G L L	G F A A P F T Q C G
Geno A	- - - - - - - - D -	- - - - - - - H - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - L - - - Q - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno D	- C Y - S - - - D -	- I - - - - E - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno F	- - - - S - - - D -	- - H - - - D - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -

Position	651 652 653 654 655 656 657 658 659 660	661 662 663 664 665 666 667 668 669 670	671 672 673 674 675 676 677 678 679 680	681 682 683 684 685 686 687 688 689 690	691 692 693 694 695 696 697 698 699 700
Reference	Y P A L M P L Y A C	I Q A K Q A F T F S	P T Y K A F L S K Q	Y L N L Y P V A R Q	R P G L C Q V F A D
Geno A	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- M - - - - - - - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - S - - - - - - -	- - - - - - - C - -	- - - - - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - S - - - - - - -	- - - - - - - C - -	- - - - - - - - - -	- - - - - - - - - -
Geno F	- - - - - - - - - -	- T - - - - - V - -	- - - - - - - C - -	- M - - - - - - - -	- - - - - - - - - -

Position	701 702 703 704 705 706 707 708 709 710	711 712 713 714 715 716 717 718 719 720	721 722 723 724 725 726 727 728 729 730	731 732 733 734 735 736 737 738 739 740	741 742 743 744 745 746 747 748 749 750
Reference	A T P T G W G L A I	G H Q R L R G T F V	S P L P I H T A E L	L A A C F A R S R S	G A K L I G T D N S
Geno A	- - - - - - - - - I	- - - - - - - - - -	A - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno B	- - - - - - - - - I	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - I	- - - - - - - - - -	A - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - - V M	- - - - - - - - - L	A - - - - - - - - -	- - - - - - - - - -	- - N I L - - - - -
Geno F	- - - - - - - - - I	- - - - - - - - - -	A - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -

Position	751 752 753 754 755 756 757 758 759 760	761 762 763 764 765 766 767 768 769 770	771 772 773 774 775 776 777 778 779 780	781 782 783 784 785 786 787 788 789 790	791 792 793 794 795 796 797 798 799 800
Reference	V V L S R K Y T S F	P W L L G C A A N W	I L R G T S F V Y V	P S A L N P A D D P	S R G R L G L Y R P
Geno A	- - - - - - - - - -	- - - - - - T - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - S - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - S - -
Geno F	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -

➔ *continues on next page*

HBV Pol Amino Acid Alignment (page 5/5)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Terminal protein | Spacer | Reverse transcriptase | RNAse H

Position	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843
Reference	L	L	R	L	L	Y	R	P	T	T	G	R	T	S	L	Y	A	D	S	P	S	V	P	S	H	L	P	D	R	V	H	F	A	S	P	L	H	V	A	W	K	P	P
Geno A	-	-	-	-	P	F	Q	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	H	-	P	F	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	P	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno F	-	-	-	-	P	F	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

HBV Large S Protein Amino Acid Alignment (page 1/2)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Pre-S2   Surface antigen

Position	1–10	11–20	21–30	31–40	41–50
Reference	MGGWSSKPRK	GMGTNLSVPN	PLGFFGDHQL	DPLFRANSEN	PDWDLNPHKD
Geno A	----------	----------	----------	--A-G---N-	----F--I--
Geno B	----------	----------	----------	--A-K-----	----------
Geno C	---------Q	G---------	----------	--A-G---N-	----F--N--
Geno D	--........	...Q---TS-	----------	--A----TA-	----F--N--
Geno E	--.L-WTVPL	EW-K-H-TT-	----------	--A----TR-	----H--N--
Geno F	--APLSTT-R	---Q------	----------	--------SS	----F-KN--

Position	51–60	61–70	71–80	81–90	91–100
Reference	NWPDANKVGV	GAFGPGFTPP	HGGLLGWSPQ	AQGRLTTVPA	APPPASTNRQ
Geno A	H--A--Q---	------L---	---I------	---I----ST	I---------
Geno B	----------	----------	----------	---L------	----------
Geno C	H--E--Q--A	----------	----------	---I------	----------
Geno D	T--------A	----L-----	----------	---I-Q-L--	N---------
Geno E	H-TE------	----------	----------	---M-K-L--	D---------
Geno F	---M------	-GY-------	----------	---V---L--	D--------R

Position	101–110	111–120	121–130	131–140	141–150
Reference	SGRQPTPLSP	PLRDTHPQAM	QWNSTTFHQT	LQDPRVRALY	LPAGGSSSGI
Geno A	-------I--	----S-----	-----A---A	-------G--	----------
Geno B	----------	----------	----------	----------	----------
Geno C	-------I--	----S-----	---------A	-L-----G--	----------
Geno D	----------	---N------	----------	-------G--	----------
Geno E	-------IT-	----------	---------A	-------G--	----------
Geno F	---K---V--	----------	-----Q---A	-L--------	--------E-

Position	151–160	161–170	171–180	181–190	191–200
Reference	VSPAQNTVSA	ISSILSTTGD	PVPNMENISS	GLLGPLLVLQ	AGFFLLTKIL
Geno A	-N--P-IA-H	----SAR---	--T-----T-	-F--------	-------R--
Geno B	----------	------K---	--------A-	----------	----------
Geno C	-N-VPT-A-P	----F-R---	-A-----TT-	-F--------	-------R--
Geno D	-N-VPT-A-P	----F-RI--	-AL-----T-	-F--------	-------R--
Geno E	-N-VPT-A-L	----F-RI--	-A----S-T-	-F--------	----------
Geno F	QN--PTIA-L	T---F-K--G	-AM--D--T-	----------	-VC-------

➔ *continues on next page*

HBV Large S Protein Amino Acid Alignment (page 2/2)

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Pre-S2 Surface antigen

Position	201 202 203 204 205 206 207 208 209 210	211 212 213 214 215 216 217 218 219 220	221 222 223 224 225 226 227 228 229 230	231 232 233 234 235 236 237 238 239 240	241 242 243 244 245 246 247 248 249 250
Reference	T I P Q S L D S W W	T S L S F L G G T P	V C L G Q N S Q S Q	I S S H S P T C C P	P I C P G Y R W M Y
Geno A	- - - - - - - - - -	- - - - - - - - S -	- - - - - - - - - -	P T - N - - - - S -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - - - A -	T - P - - - - - - P	T - N - - - - S - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - - T	- - - - - - - - - -	P T - N - - - - S -	- T - - - - - - - -
Geno E	- - - - - - - - - -	- - - - - - - - A -	- - - - - - - - - -	P T - N - - - - S -	- - - - - - - - - -
Geno F	- - - - - - - - - -	- - - - - - - - L -	G - P - - - - - - P	T - N - L - - S - -	- T - - - - - - - -

Position	251 252 253 254 255 256 257 258 259 260	261 262 263 264 265 266 267 268 269 270	271 272 273 274 275 276 277 278 279 280	281 282 283 284 285 286 287 288 289 290	291 292 293 294 295 296 297 298 299 300
Reference	L R R F I I F L C I	L L L C L I F L L V	L L D Y Q G M L P V	C P L I P G S S T T	S T G P C K T C T T
Geno A	- - - - - - - - F -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - T - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - F -	- - - - - - - - - -	- - - - - - - - - -	- - - L - - T - - -	- - - - - - - - - I
Geno D	- - - - - - - - F -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno E	- - - - - - - - F -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno F	- - - - - - - - F -	- - - - - - - - - -	- - - - - - - - - -	- - - L - - - T - -	- - - - - - - - - -

Position	301 302 303 304 305 306 307 308 309 310	311 312 313 314 315 316 317 318 319 320	321 322 323 324 325 326 327 328 329 330	331 332 333 334 335 336 337 338 339 340	341 342 343 344 345 346 347 348 349 350
Reference	P A Q G T S M F P S	C C C T K P T D G N	C T C I P I P S S W	A F A K F L W E W A	S V R F S W L S L L
Geno A	- - - - N - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - Y - - - - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - Y - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - S - - -	- - - - - - - - - -	- - - R - - - - - -	- - - - - - - - - -
Geno D	- - - - - - - Y - -	- - - - - - S - - -	- - - - - - - - - -	- - G - - - - - - -	- A - - - - - - - -
Geno E	L - - - - - - - - -	- - - S - - S - - -	- - - - - - - - - -	- - G - - - - - - -	- A - - - - - - - -
Geno F	L - - - - - - - - -	- - - S - - S - - -	- - - - - - - - - -	- L G - Y - - - - -	- A - - - - - - - -

Position	351 352 353 354 355 356 357 358 359 360	361 362 363 364 365 366 367 368 369 370	371 372 373 374 375 376 377 378 379 380	381 382 383 384 385 386 387 388 389 390	391 392 393 394 395 396 397 398 399 400
Reference	V P F F V Q W F V G L	S P T V W L S V I W	M M W Y W G P S L Y	N I L S P F M P L L	P I F F C L W V Y I
Geno A	- - - - - - - - - -	- - - - - - - A - -	- - - - - - - - - -	S - V - - - I - - -	- - - - - - - - - -
Geno B	- - - - - - - - - -	- - - - - - - - - -	- - - F - - - - - -	- - - - - - - - - -	- - - - - - - - - -
Geno C	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - L - - -	- - - - - - - - - -
Geno D	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	S - - - - - L - - -	- - - - - - - - - -
Geno E	- - - - - - - A - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - I - - -	- - - - - - - - - -
Geno F	- Q - - - - C - - -	- - - - - - L - - -	- I - - - - - - - C	S - - - - - I - - -	- - - C Y - - - S -

HBV X Protein Amino Acid Alignment

Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Reference	M	A	A	R	L	C	C	Q	L	D	P	A	R	D	V	L	R	L	R	P	V	G	A	E	S	R	G	R	P	L	P	G	P	L	G	A	L	P	P	A	S	P	P	V	V	P	T	D	H	G
Geno A	-	-	-	-	-	Y	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	T	-	S	S	P	-	-	S	A	-	-	A	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	v	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	S	-	-	F	-	T	-	-	S	P	-	-	S	A	-	-	A	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	S	-	-	-	-	T	-	S	S	P	-	-	S	A	-	S	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	C	-	-	-	V	S	-	S	-	-	D	-	S	S	P	-	-	S	A	-	-	A	-	-	-
Geno F	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	S	-	S	-	-	-	V	-	-	P	-	-	S	A	-	-	A	-	D	-

Position	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Reference	A	H	L	S	L	R	G	L	P	V	C	A	F	S	S	A	G	P	C	A	L	R	F	T	S	A	R	R	M	E	T	T	V	N	A	H	R	N	L	P	K	V	L	H	K	R	T	L	G	L
Geno A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	Q	I	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	V	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	F	-	-	-	-	-	-	-	-	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	I	-	-	-	-	-	-	-	-	-	-	-	-
Geno F	S	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	S	-	-	T	-	-	-	-	-	-	-	-

Position	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Reference	S	A	M	S	T	T	D	L	E	A	Y	F	K	D	C	V	F	T	E	W	E	E	L	G	E	E	I	R	L	M	I	F	V	L	G	G	C	R	H	K	L	V	R	S	P	A	P	C	N	F
Geno A	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	D	-	-	-	-	-	-	-	-	-	-	K	V	-	-	-	-	-	-	-	-	-	-	-	C	A	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	V	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-
Geno D	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	K	D	-	-	-	-	-	-	-	-	-	-	K	V	-	-	-	-	-	-	-	-	-	-	-	C	A	-	-	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	K	D	-	-	-	-	-	-	-	-	-	-	K	V	-	-	-	-	-	-	-	-	-	-	-	C	A	-	-	-	-	-	-
Geno F	-	G	R	-	M	-	W	I	-	E	-	I	-	-	-	-	-	K	D	-	-	-	-	-	-	-	-	-	-	K	V	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-

Position	151	152	153	154
Reference	F	T	S	A
Geno A	-	-	-	-
Geno B	-	-	-	-
Geno C	-	-	-	-
Geno D	-	-	-	-
Geno E	-	-	-	-
Geno F	-	-	-	-

HBV Pre-Core Protein Amino Acid Alignment (first 29 residues only)

The remaining residues (30 onwards) are represented in the alignment for core sequence (see page 11)
Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Reference	M	Q	L	F	H	L	C	L	I	I	S	C	S	C	P	T	V	Q	A	S	K	L	C	L	G	W	L	W	G
Geno A	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno F	-	-	-	-	-	-	-	-	-	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

HBV Core Protein Amino Acid Alignment

The pre-core protein incorporates an additional 29 residues upstream of this sequence (see sequences represented on page 10).
Reference sequence is from Liu WC, et al., *Hepatol. Int.* 2016;10:147-57.
Consensus sequences are derived from all available sequences at (https://hbvdb.ibcp.fr); see methods (page 1).
'-' designates a residue the same as that within the reference sequence.
'.' designates missing data
Residues that differ from the reference sequence are specified.
The sequence is numbered consecutively from 1 and divided into blocks of ten residues (dashed lines) for ease of navigation.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Reference	M	D	I	D	T	Y	K	E	F	G	A	S	V	E	L	L	S	F	L	P	S	D	F	F	P	S	I	R	D	L	L	D	T	A	T	A	L	Y	R	E	A	L	E	S	P	E	H	C	S	P
Geno A	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-
Geno F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	T	-	-

Position	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Reference	H	H	T	A	L	R	Q	A	I	V	C	W	G	E	L	M	N	L	A	T	W	V	G	S	N	L	E	D	P	A	S	R	E	L	V	V	S	Y	V	N	V	N	M	G	L	K	I	R	Q	L
Geno A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	T	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	F	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	F	-	-	-	-
Geno F	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	S	-	-	-	N	-	-	-	-	-	-	A	-	-	-	-	-	N	-	-	T	-	-	-	-	-	-	-	-	-	-

Position	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Reference	L	W	F	H	I	S	C	L	T	F	G	R	E	T	V	L	E	Y	L	V	S	F	G	V	W	I	R	T	P	P	A	Y	R	P	A	N	A	P	I	L	S	T	L	P	E	T	T	V	V	R
Geno A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
Geno F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Position	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183
Reference	R	R	G	R	S	P	R	R	R	T	P	S	P	R	R	R	R	S	Q	S	P	R	R	R	R	S	Q	S	R	E	S	K	C
Geno A	-	-	D	.	R	G	-	.	.	.	.	-	-	-	-	-	T	P	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Geno E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	A	-	-
Geno F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	A	-	-