



- (51) **International Patent Classification:**
G06F 7/00 (2006.01)
- (21) **International Application Number:**
PCT/US2018/018092
- (22) **International Filing Date:**
14 February 2018 (14.02.2018)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PCT/US2017/017842
14 February 2017 (14.02.2017) US
PCT/US2017/041591
11 July 2017 (11.07.2017) US
- (71) **Applicant:** GENOMSYS SA [CH/CH]; Chemin De La Raye 13, CH-1024 Ecublens (CH).
- (72) **Inventor; and**
- (71) **Applicant:** BALUCH, Mohamed, Khoso [US/US]; 4439 Woodsedge Ct, Chantilly, VA 20151 (US).
- (72) **Inventors:** ALBERTI, Claudio; Chemin Des Esserts 1, CH-1213 Petit-Lancy (CH). ZOIA, Giorgio; Chemin Des Croix-rouges 10, CH-1007 Lausanne (CH). RENZI,
- (74) **Agent:** ETTEHADIEH, Aslan; Aslan Law, P.C., 7804 Holmes Run Drive, Falls Church, VA 22042 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** METHOD AND APPARATUS FOR THE COMPACT REPRESENTATION OF BIOINFORMATICS DATA USING MULTIPLE GENOMIC DESCRIPTORS

(57) **Abstract:** Method and apparatus for the compression of genome sequence data produced by genome sequencing machines. Sequence reads are coded by aligning them with respect to pre-existing or constructed reference sequences, the coding process is composed of a classification of the reads into data classes followed by the coding of each class in terms of a multiplicity of descriptors blocks. Specific source models and entropy coders are used for each data class in which the data is partitioned, and each associated descriptor block.

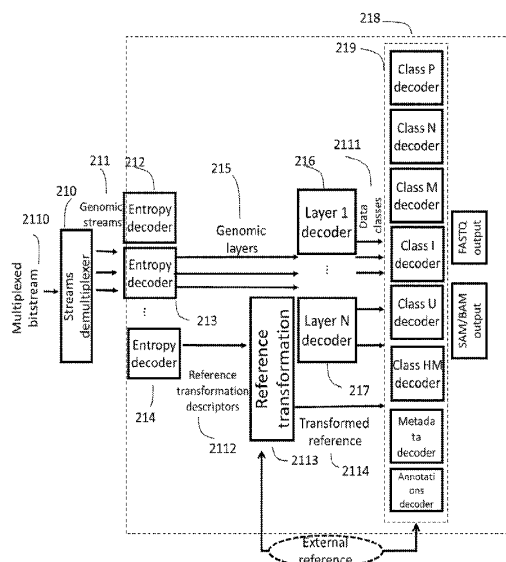


Figure 21

Published:

— with international search report (Art. 21(3))