



(51) International Patent Classification:
C12Q 1/68 (2006.01) *G06F 7/00* (2006.01)

(21) International Application Number:
PCT/US2010/022912

(22) International Filing Date:
2 February 2010 (02.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/149,665 3 February 2009 (03.02.2009) US

(71) Applicant (for all designated States except US): **COMPLETE GENOMICS, INC.** [US/US]; 2071 Stierlin Court, Mountain View, CA 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HALPERN, Aaron, L.** [US/US]; 940 Walnut Street, San Carlos, CA 94070 (US). **NAZARENKO, Igor** [US/US]; 660 Gail Avenue, No. A4, Sunnyvale, CA 94086 (US).

(74) Agents: **DEVORE, Dianna, L.** et al.; Virtual Law Partners, LLP, P.O.Box 1329, Mountain View, CA 94042 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM,

[Continued on next page]

(54) Title: INDEXING A REFERENCE SEQUENCE FOR OLIGOMER SEQUENCE MAPPING

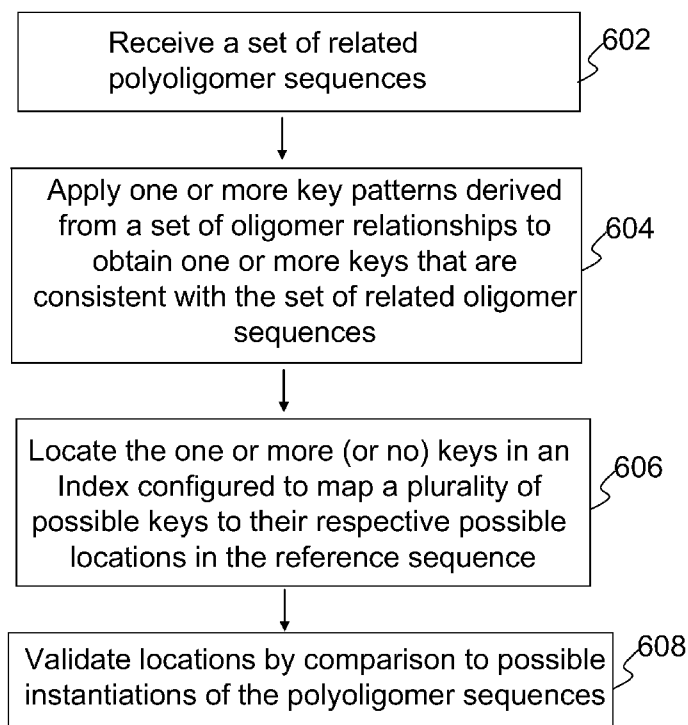


FIG. 6

(57) Abstract: Generating an index includes receiving a reference sequence and applying one or more key patterns to the reference sequence to obtain a plurality of keys in the index. Based on its inputs, the key generator generates one or more keys, which are strings of base sequences, compressed representations of base sequences or any other appropriate representations of base sequences that are suitable for computer processing. Each of the one or more key patterns is derived based on a corresponding set of oligomer sequence relationships of a plurality of oligomer sequences that are expected to be generated from the reference, and the keys correspond to a plurality of candidate and/or validated locations in the reference sequence.





TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, (88) Date of publication of the international search report:
ML, MR, NE, SN, TD, TG). 30 September 2010

Published:

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/22912

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68; G06F 7/00 (2010.01)

USPC - 435/6; 707/741;

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC- C12Q 1/68; G06F 7/00 (2010.01);

USPC- 435/6; 707/741

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC- 702/19; 706/45 706/46 707/741 707/758 707/801 707/802 707/E17.039 707/E17.002 707/E17.044; 382/209; 715/256;
Patents and NPL

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWest (US Pat, PgPub, EPO, JPO: classification, keyword), GoogleScholar;
search terms: key, pattern, token, sequence, ma, dna, oligomer, predict, generate, distance, location, index, encode, compress

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,577,249 A (CALIFANO) 19 November 1996 (19.11.1996), col 2, ln 52-67; col 5, ln 25-60;	1-8, 11-24
-	col 6, ln 16-19; col 7, ln 8 to col 8, ln 38; col 9, ln 2-40; col 10, ln 10 to col 12, ln 61; col 15, ln 4-	-----
Y	60; col 16, ln 23-67	9, 10
Y	US 2005/0187916 A1 (LEVIN et al.) 25 August 2005 (25.08.2005), para [0049]-[0051], [0092]-[0094], [0208], [0209]	9, 10
A	US 2008/0256070 A1 (INGLIS) 16 October 2008 (16.10.2008), entire document	1-24
A	US 2008/0040045 A1 (SELIFONOV et al.) 14 February 2008 (14.02.2008), entire document	1-24
A	US 2007/0225918 A1 (SAYOOD et al.) 27 September 2007 (27.09.2007), entire document	1-24
A	US 6,055,526 A (AMBROZIAK) 25 April 2000 (25.04.2000), entire document	1-24

☐ Further documents are listed in the continuation of Box C.


* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 June 2010 (13.06.2010)

Date of mailing of the international search report

23 JUN 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774