

(51) International Patent Classification:
G01N 33/50 (2006.01)(21) International Application Number:
PCT/US2009/047836(22) International Filing Date:
18 June 2009 (18.06.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/074,150 20 June 2008 (20.06.2008) US(71) Applicants (for all designated States except US): **EUREKA GENOMICS CORPORATION** [US/US]; 750 Alfred Nobel Drive, Suite 108, Hercules, California 94547-1805 (US). **UNIVERSITY OF HOUSTON SYSTEM** [US/US]; Office of Intellectual Property Management, Attention: Mark Yzaguirre, 316 East Cullen, Houston, Texas 77204-2015 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **FOFANOV, Yuriy** [US/US]; 19 Nila Grove Court, The Woodlands, Texas 77385 (US). **KOSHINSKY, Heather** [CA/US]; 2033 Carquinez Avenue, El Cerrito, California 94530 (US). **FOFANOV, Viacheslav** [US/US]; 2232 Bellefontaine Drive, #3, Houston, Texas 77030 (US). **FENG, Chen** [US/US]; 6406 Crystal Point, Missouri City, Texas 77459 (US).(74) Agents: **WORRALL, Timothy A.** et al.; 370 Seventeenth Street, Suite 4700, Denver, Colorado 80202 (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, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

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

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR SEQUENCING DATA SAMPLES

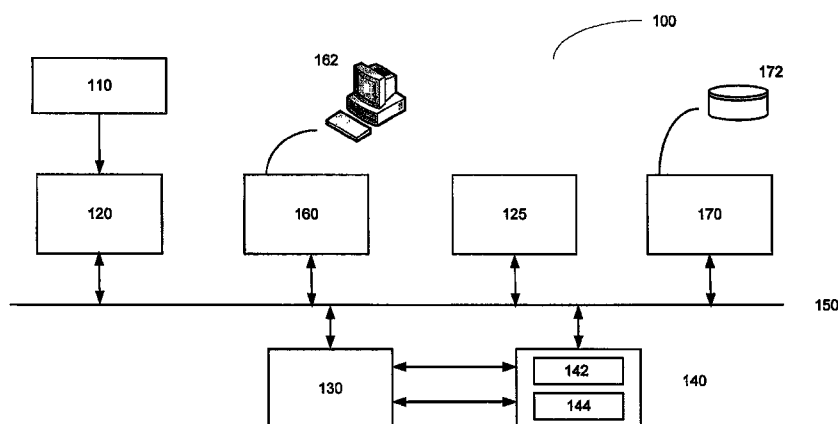


FIG. 1

(57) **Abstract:** A method for identifying non-host nucleic acid sequence using sequence data. The method of identifying non-host nucleic acid can include sequencing a sample into sequences and associating the sequences with a host genome and then exclude any sequences that are associated with the host genome. The method can then associate the sequences with any known genomes and exclude any sequences that are associated with any known genome. The remaining sequences can be used as seed sequences to assemble a non-host nucleic acid.



— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(88) Date of publication of the international search report:
25 February 2010

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 09/47836

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G01N 33/50 (2009.01)

USPC - 702/20

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 (8) G01N 33/50; (2009.01)

USPC- 702/20;

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

IPC (8) G01N33/48 (2009.01)

USPC- 435/6

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST(PGPB,USPT,USOC,EPAB,JPAB); Google Patents; Google Scholar

nucleic acid sequencing, sequence, data analysis, suffix, table, tree, hash table, sorted array, algorithm, mismatch, Koshinsky, Fofanov, assembly, data structure, reference, identify,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	(NING et al.) SSAHA: A Fast Search Method for Large DNA Databases Genome Res. 2001 11: 1725-1729. (pg 1721 col 1 para 3) (pg 1725 col 2 para 3) (pg 1726 col 1 para 1) (pg 1726 col 2 para 2,3) (pg 1727 col 1 para 2) (pg 1727 col 2 para 2) (pg 1728 col 2 para 3-6) (pg 1729 col 1 para 2).	1-18
A	US 2006/0286566 A1 (LAPIDUS et al.) 21 December 2006 (21.12.2006) (abstract) (para [0007]-[0015], [0053]).	1, 3-12, and 14-18

☐ 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

16 December 2009 (16.12.2009)

Date of mailing of the international search report

04 JAN 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