

(51) International Patent Classification:
G06F 19/10 (2011.01)

(21) International Application Number:

PCT/US2014/030014

(22) International Filing Date:

15 March 2014 (15.03.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/786,398 15 March 2013 (15.03.2013) US

(71) Applicant: **ANCESTRY.COM DNA, LLC** [US/US]; 360 W 4800 N, Provo, UT 84604 (US).(72) Inventors: **BARBER, Mathew**; Ancestry.com DNA, LLC, 360 W 4800 N, Provo, UT 84604 (US). **CURTIS, Ross**; Ancestry.com DNA, LLC, 360 W 4800 N, Provo, UT 84604 (US). **BALL, Catherine**; Ancestry.com DNA, LLC, 360 W 4800 N, Provo, UT 84604 (US). **CHAHINE, Kenneth, Gregory**; Ancestry.com DNA, LLC, 360 W 4800 N, Provo, UT 84604 (US).(74) Agents: **BROWNSTONE, Daniel, R.** et al.; Fenwick & West LLP, 555 California St. FL 12, San Francisco, CA 94104 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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, 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, 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,

[Continued on next page]

(54) Title: FAMILY NETWORKS

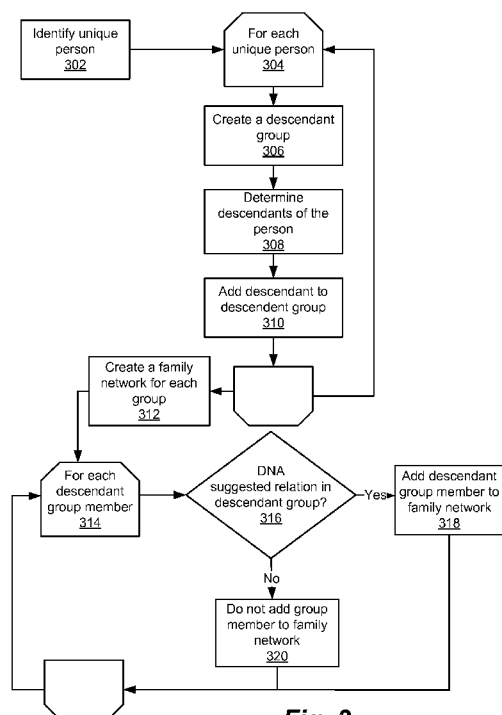


Fig. 3

(57) Abstract: Described embodiments enable identification of family networks using combinations of DNA analysis and genealogical information. Genealogical data is provided by users of a genealogical research service or collected from other sources and used to create family trees for each user. DNA samples are also received from the users. By analyzing the DNA samples, potential genetic relationships can be identified between some users. Once these DNA-suggested relationships have been identified, common ancestors can be sought in the respective trees of the potentially related users. Where these common ancestors exist, an inference is drawn that the DNA-suggested relationship accurately represents a familial overlap between the individuals in question. People descended from the same common ancestor are each members of a family network. Members of a family network not in a user's tree may be identified for the user, enabling the user to discover additional ancestors that might otherwise have remained unknown.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, —
KM, ML, MR, NE, SN, TD, TG).

*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))*

Published:

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

FAMILY NETWORKS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application 61/786,398, filed on March 15, 2013, which is incorporated by reference in its entirety.

BACKGROUND

Field

[0002] This disclosure relates generally to computer software for identification of family relationships based on combinations of DNA matching and genealogical records.

Description of Related Art

[0003] Genealogical research is typically undertaken by individuals interested in learning more about their family history. Generally, researchers build their family trees by entering information about known ancestors, including, for example, birth and death dates and locations, spouses, offspring, and the like. Software designed for genealogical research is often used for this purpose, and may be used in a standalone fashion, or via a networked implementation. Some genealogical research services offer suggestions to researchers about additional information that may be available about relatives already present in their family trees such as census information, immigration records, etc.

[0004] Services also exist for extracting and characterizing DNA samples from individuals. Some services identify similar DNA segments between customers and suggest to the customers having those similar segments that they may be related.

SUMMARY

[0005] Described embodiments enable identification of family networks using combinations of DNA analysis and genealogical information. Genealogical information is provided by users of a genealogical research service or collected from other sources and used to create family trees for each of the users. DNA samples are also received from the users and analyzed. By comparing the results of the DNA analysis, potential genetic relationships can be identified between some users. Once these DNA-suggested relationships have been identified, common ancestors can be sought in the respective family trees of the potentially related users. Where these common ancestors exist, an inference is drawn that the DNA-suggested relationship accurately represents a familial overlap between the individuals in question.

[0006] People descended from a common ancestor are each members of a family network, though no single genealogical tree compiled by a single user may yet include all of the members of the network. In various embodiments, members of a family network not in a user's tree may be identified for the user. In this way, a user can discover additional ancestors that might otherwise have remained a mystery.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Fig. 1 is a block diagram illustrating components of a system for identifying family networks in accordance with one embodiment.

[0008] Fig. 2 illustrates a portion of a family tree in accordance with one embodiment.

[0009] Fig. 3 is a flowchart illustrating a method for creating family networks in accordance with one embodiment.

[0010] Fig. 4 illustrates a method for creating family networks in accordance with one embodiment.

[0011] Fig. 5 illustrates a portion of a family tree in accordance with one embodiment.

[0012] Fig. 6 illustrates a portion of a family tree in accordance with one embodiment.

[0013] Fig. 7 illustrates a method for collapsing nodes in accordance with one embodiment.

[0014] Fig. 8 illustrates an example of a user interface illustrating birth locations of a customer's DNA suggested relations in accordance with one embodiment.

DETAILED DESCRIPTION

[0015] Fig. 1 illustrates an example system 100 for identifying family networks in accordance with one embodiment. System 100 includes a trees most recent common ancestor (MRCA) service 102, customer DNA data store 104, customer trees data store 106, customer account data store 108, hints engine 110, a DNA relationship engine 114, and tree editor 116. Fig. 1 also illustrates a DNA extraction service 112. Each of these elements is described further below.

[0016] For purposes of clarity within this description, we assume that system 100 is administered by or on behalf of a company providing genealogical research services to its customers, though many other use cases will be apparent from the disclosure. One example of such a company is Ancestry.com, of Provo, Utah. Services may be provided to customers via the web, in person, by telephone, by mail, or various combinations of the above.

[0017] Customers of system 100 (which we also refer to interchangeably as "users") have accounts with system 100, and account data in various embodiments is stored in customer account data store 108. Customers may create family trees containing genealogical data known to the customer or obtained by the customer from the genealogical research service. For example, tree editor 116 in various embodiments includes a web interface through which users can enter or upload genealogical data using their computers or mobile devices. System 100 stores tree data for a customer in customer tree data store 106. In various embodiments, trees include records describing individuals that are thought by the curator of the tree (e.g., the user) to be related. Each record may include, for example, an individual's name, date of birth, date of death, birth and death locations, places lived,

education, and other information known about the individual. In addition, the record describes connections between the individual and other people in the tree, for example by describing the individual's parents, siblings, spouses and children. In various embodiments, these connections are links or pointers to other individual records in the same tree. Records, which are stored in customer tree data store 106, may be added, deleted or modified as desired by curators of the tree. For example, Fig. 2 illustrates a portion of a family tree 200 of Abraham Lincoln (1809-1865). Duplicate records for any particular individual may exist in customer trees data store 106 as a consequence of different users creating independent records for the same individual in their respective family trees—for example, two users of system 100 may be related to each other as third cousins, and each may have a family tree that includes their great-great-grandfather—an ancestor common to both cousins. As described below, these multiple records describing a single individual, when combined with DNA-suggested relationships, lead to the discovery and expansion of family networks.

[0018] In addition to providing genealogical data to populate family trees, customers of system 100 may also provide DNA samples for analysis of their genetic data. In one embodiment, a customer obtains a sample collection kit associated with his account in customer account data store 108. The customer uses the sample collection kit to provide a sample, e.g., saliva, from which genetic data can be reliably extracted according to conventional methods. DNA extraction service 112 receives the customer sample and genotypes the genetic data, for example by extracting the DNA from the sample and identifying values of single nucleotide polymorphisms (SNPs) present within the DNA. In one embodiment, genotyping takes place at a large number, e.g., 700,000 SNP locations in the genome. In various embodiments, inferences about the customer may be made using the SNP data, for example including a prediction of the customer's ethnic background and the geographic migration patterns of the customer's ancestors. System 100 receives the genetic data from DNA extraction service 112 and stores the genetic data in customer DNA data store 104 along with an association to the customer account in customer account data store 108.

[0019] Note that in various embodiments, depending on the information a particular customer has chosen to provide, the customer may have family tree data stored in customer tree data store 106, genetic data stored in customer DNA data store 104, or both.

[0020] DNA relationship engine 114 analyzes the genetic data contributed by customers and stored in DNA data store 104 to identify potential genetic relationships between the customers. In one embodiment, the GERMLINE algorithm is used to identify shared segments of genotype data between users. The GERMLINE algorithm is described in Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I (2008) *Whole population, genomewide mapping of hidden relatedness*, Genome Research, and is incorporated by reference herein in its entirety.

[0021] Each potential genetic relationship identified by DNA relationship engine 114 is noted in the customer account data for each of the potentially matching customers if the potential genetic relationship is significant. The threshold level for a significant genetic relationship is set depending on the preference of the implementer. In various embodiments, a potential genetic relationship between two customers is identified by DNA relationship engine 114 as significant if a significant similarity exists at the same part of their genomes, for example, in one embodiment at least $5 * 10^6$ continuous base pairs are identical between two individuals. Other thresholds may be set according to the preference of the implementer. We refer in this disclosure to two or more customers for whom an identification of a significant genetic relationship has been made as "DNA suggested relations." In various embodiments, a confidence score indicating the degree of genetic similarity between the DNA suggested relations is calculated and stored in the customer account data.

Family Network Identification

[0022] A group of people, each descended from a particular common ancestor for whom a record exists in customer tree data store 106 (and including the ancestor's record) constitutes a family network. In some embodiments, a family network is further limited to

include only those people who descend according to customer tree data from a common ancestor and who are also DNA suggested relations with at least one other member of the tree based on data stored in DNA data store 104.

[0023] For any group of two or more people having records in tree data store 106, we refer to their closest common ancestor(s) from a generational perspective as their “most recent common ancestor(s)” or MRCA. For example, referring again to the family tree in Fig. 2, the most recent common ancestors of Jessie Harlan Lincoln and Thomas Tad Lincoln are Abraham Lincoln and his wife Mary Ann Todd. More formally, in one embodiment for any two people, their MRCAs are the people from whom they are directly descended and for whom no other person from whom they are directly descended is generationally closer.

[0024] In one embodiment, and referring now to Fig. 3, family networks are identified as follows. Trees MRCA service 102 examines each record in customer tree data 106 to identify 302 each unique person. That is, a particular person may be named in multiple trees as described above—for example if many different users have an instance of the same person in each of their trees—and therefore be represented by multiple records, but be in fact a unique individual. Multiple instances of a single individual in various embodiments are identified by comparing names, birth dates, death dates, places of residence and other bibliographic data to establish a likelihood that any two or more records in fact describe the same individual. For each unique person identified 304, MRCA service 102 creates 306 a descendant group and identifies 308 all known descendants of that unique individual across all customer trees in tree data store 106. Each descendant is then added 310 to the descendant group for the unique individual. MRCA service 102 repeats the process for each unique individual.

[0025] At least one family network is created 312 for each descendant group. In one embodiment, at the conclusion of steps 302-310 each descendant group anchored by the unique individual, e.g., the unique common ancestor, is designated as a family network. In another embodiment, and still referring to Fig. 3, each descendent group member in a

descendant group is evaluated 314 to determine whether 316 that descendent group member has a DNA-suggested relation who is also in the descendant group. If so, the descendant group member 318 is added to the family network anchored by the unique individual. Otherwise, the descendent group member is not added 320 to the family network. In this embodiment, a family network includes only a common ancestor and individuals who are shown to be related via both genealogical and genetic data and is more likely to exclude tree members who were added incorrectly, for example through errors in the curation of the genealogical data.

[0026] In various alternative embodiments, family networks are created by using DNA-suggested relationships to search for most recent common ancestors. Fig. 4 illustrates a method for creating family networks in accordance with one such alternative embodiment. Genetic data is obtained, for example by receipt 402 of a customer saliva sample, and DNA is extracted 404 from the sample by DNA extraction service 112 as described above. Genetic data is obtained 406 from the DNA and stored in customer DNA data store 104. DNA relationship engine 114 uses genetic data 406 across multiple customer DNA data store 104 to identify 408 DNA-suggested relations as described above.

[0027] For each of the users in a pair of DNA-suggested relations, hints engine 110 searches the users' respective trees in tree data store 106 to locate 410 within the trees one or more potential common ancestors. As described above with respect to Fig. 3, common ancestors with instances in both trees may be identified by comparing names and other bibliographic data to establish a likelihood that any two individuals are in fact the same person. Hints engine 110 then creates a hint including indicia of the records from each tree for the potential match. Indicia may include, for example, a record identifier for the record.

[0028] For example, referring to Fig. 5, consider a family tree 500 that includes four people – D 510, D's children F 504 and H 508, F's spouse G 506, and A 502, the child of F and G. A second family tree 600 is illustrated in Fig. 6. This second tree also includes person D

510 and person H 508, and additionally includes D's spouse E 602; H's spouse J 604; and B 606, the child of H and J.

[0029] As can be seen from the two trees above when viewed together, D 510 is a common ancestor of A 502 and B 606, though this is not apparent from either of the trees when considered individually.

[0030] Assume that both tree 500 and tree 600 are stored in tree data store 106. Assume also that individuals A 502 and B 606 have submitted DNA samples and been identified as DNA-suggested relations.

[0031] Hints engine 110 identifies A and B as DNA-suggested relations and searches their respective trees 500 and 600 for instances of common ancestors as described above with respect to Fig. 4. Since individual D 510 is in both tree 500 and tree 600—that is, D is a common ancestor, hints engine 110 creates a hint with indicia of both records. For example:

[0032] Hint #1: (D, Tree1); (D', Tree2).

[0033] The notation D and D' illustrates for purposes of this disclosure two distinct records in trees data store 106 that each refer to the same unique individual.

[0034] If individual E (married, in this example, to D, though not illustrated in Fig. 5) also exists in both trees, hints engine 110 identifies a second hint:

[0035] Hint #2: (E, Tree1); (E', Tree2).

[0036] Referring now to Fig. 7, each hint for a most recent common ancestor (MRCA) can be considered a node for a common ancestor. Hints engine 110 therefore can create 702 two nodes in the above example:

[0037] Node #1: (D, Tree1; D', Tree2).

[0038] Node #2: (E, Tree1; E', Tree2).

[0039] Once hints engine 110 has created all possible nodes across all trees for all individuals with DNA-suggested relationships, each node can be further collapsed 704. For

example, beginning with Node #1, the record D, Tree1 may exist in another node, such as one associated with another DNA-suggested relation of user A:

[0040] Node #3: (D, Tree1, D', Tree3).

[0041] Hints engine 110 then merges the nodes:

[0042] Node #1: (D, Tree1; D', Tree2; D'', Tree3).

[0043] Thus, for each node 704, hints engine 110 selects 706 an identifier in the node and determines 708 whether the identifier is present in another node. If so, the nodes are merged 710. By iterating over all nodes, a concatenated set of nodes is created. Within each node, all of the records point to different instances in different trees of the same person, i.e. the common ancestor of the concatenated set of DNA-suggested relations. The common ancestor combined with the DNA-suggested relations together constitute a family network. In one embodiment, hints engine 110 updates the records for each of the members of the family network to reflect this finding. In an alternative embodiment, hints engine 110 recommends 412 (Fig. 4) the addition of the family network members to each user, leaving to the user the option of adding one or more of the recommended relatives to the user's tree.

[0044] In various embodiments, the analysis described above is run once at set up time, and then periodically thereafter. In addition, or alternatively, the analysis is repeated when one or more new DNA samples are received by customer DNA data store 104.

[0045] When one or more customers are identified as part of a family network, in various embodiments system 100 notifies the customers of the finding. This may include, for example, introducing DNA-suggested relations within the family network to one another; or identifying ancestors earlier than the most recent common ancestor who may be missing from trees of some members of the family network.

[0046] One or more customers identified by DNA relationship engine 114 as having a sufficient genetic similarity to indicate a possible relationship may not yet have a family tree in customer tree data store 106 complete enough to see common ancestors. For example, the

customer may not have any tree at all (and so no common ancestor will be seen between the customer and the other users in the database).

[0047] An opportunity exists to enable the customer without a sufficient tree for explicit identification of common ancestors to begin building their tree using the identified common ancestor(s) from the other customers' trees. For example, referring again to Fig. 2, if a customer is identified as a DNA-suggested relation of Mary Lincoln (1869-1938), Abraham J Lincoln II (1873-1890) and Jessie Harlan Lincoln (1875-1948), each of whom has a known common ancestor in Robert Todd Lincoln (1843-1926), then hints engine 110 can suggest that Robert Todd Lincoln should be in the new customer's family tree.

[0048] In some embodiments a confidence score is used to determine whether or not the number of matches between a customer and a family network is significant. For example, a high-confidence hint for a customer would result if a customer had eight DNA-suggested relations in a family network of size eight. One method for estimating a confidence level is to create a score based on either or both of two factors: the total number of people in each family network—e.g., five DNA-suggested relations in a family network of size six is more significant than five DNA-suggested relations in a family network of size eight; and the total number of matches a customer has—e.g., four DNA-suggested relations in a family network of size six is more significant if that customer only has 100 DNA-suggested relations rather than 10,000. By adapting to these factors, the score is broadly comparable (on approximately the same scale) across all family networks and all customers. In constructing family networks as described above, if the probability of observing the identified number of DNA-suggested relations in a family network is small, then that family network is significant, and suggests to a meaningful level of confidence that the customer is related to the family network through a mutual common ancestor.

[0049] We begin with the assumption that a customer's DNA will on occasion match some members of a random family network to whom the customer is not closely related—for example, because of false matching, relationships through another unrelated line,

relationship between members of the family network, or the like. The confidence score therefore reflects a determination that the number of matches (to a family network) identified is more significant than a random occurrence. A simple binomial sampling model or another statistical model is used (valid only under various approximately true assumptions) to calculate a coarse probability of observing at least that many of the DNA customer's DNA-suggested relations in that family network. A simple transformation of the coarse probability is the quantitative score that hints engine 110 uses to measure interest level between a customer and a family network. In one embodiment, hints engine 110 estimates a level of confidence in a family network hint using a single quantitative score.

[0050] To increase the degree of confidence in the score that other individuals should also belong to that family network, in various embodiments system 100 also considers birth locations in the tree, known surnames in the tree, and the strength of DNA matches to the descendants in the network.

[0051] In one embodiment, hints engine 110 uses DNA suggested relations to hint to a customer information to help her build her family tree stored in the customer trees data store 106. In this embodiment, hints engine 110 provides suggestions to the customer of specific places and surnames where her family might have lived.

[0052] Consider, for example, a customer with multiple distant DNA suggested relations stored in customer DNA data 104. By considering the DNA suggested relations that have trees in customer trees data store 106, hints engine 110 can identify counties, surnames, or other information that is common across a customer's DNA suggested relation's trees. To normalize these counts, hints engine 110 creates a normalization set that includes the total count of all counties, surnames, or other information of interest. In some embodiments, this information can be aggregated across all parts of the trees, or it can be limited to specific time periods of interest.

[0053] For example, consider a customer who does not have a tree. By aggregating information across m DNA suggested relations for this customer, hints engine 110 finds that

Allegheny County, Pennsylvania is found in the tree of k DNA suggested relations. In the normalization set, Allegheny County is found in L customer trees. Given n , the total number of DNA customers with data in customer DNA data store **104**, hints engine **110** can use a statistical model to determine the significance of observing k instances of Allegheny County in m matches. In one embodiment, the binomial distribution is used with parameters p and m , where p is equal to L / n . If Allegheny County is found to be significant, it is given to the customer as a hint to where that customer's family might have lived.

[0054] Hints engine **110** delivers the significant information to the customer in one embodiment in the form of tree-building hints. In one embodiment, the hints engine delivers a list of the top five surnames that the customer might expect to find in her tree. In another embodiment, and referring to Fig. 8, hints engine **110** renders a map view **800** illustrating a number of countries of birth for a customer's DNA suggested relations born during certain time periods. In another embodiment, the customer views the locations on the map to get a general feel for where their family has lived, suggesting counties in which to search for records.

[0055] In addition to the embodiments specifically described above, those of skill in the art will appreciate that the invention may additionally be practiced in other embodiments. For example, in an alternative embodiment, DNA extraction service **112** is part of system **100**.

[0056] Although this description has been provided in the context of specific embodiments, those of skill in the art will appreciate that many alternative embodiments may be inferred from the teaching provided. Furthermore, within this written description, the particular naming of the components, capitalization of terms, the attributes, data structures, or any other structural or programming aspect is not mandatory or significant unless otherwise noted, and the mechanisms that implement the described invention or its features may have different names, formats, or protocols. Further, some aspects of the system may be implemented via a combination of hardware and software or entirely in

hardware elements. Also, the particular division of functionality between the various system components described here is not mandatory; functions performed by a single module or system component may instead be performed by multiple components, and functions performed by multiple components may instead be performed by a single component. Likewise, the order in which method steps are performed is not mandatory unless otherwise noted or logically required.

[0057] Unless otherwise indicated, discussions utilizing terms such as “selecting” or “computing” or “determining” or the like refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0058] Electronic components of the described embodiments may be specially constructed for the required purposes, or may comprise one or more general-purpose computers selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, DVDs, CD-ROMs, magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, application specific integrated circuits (ASICs), or any type of non-transitory media suitable for storing electronic instructions, and each coupled to a computer system bus.

[0059] Finally, it should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure is intended to be illustrative, but not limiting, of the scope of the invention.

[0060] We claim:

1. A computer-implemented method for identifying genealogically related individuals, the method comprising:
 - receiving genetic data sampled from a first human individual;
 - identifying, by a computer, using the received genetic data, a second human individual genetically related to the first human individual;
 - identifying, by the computer, a common ancestor of the first human individual and the second human individual, the common ancestor identified by comparing individuals identified in a first set of genealogical data that are ancestors to the first human individual with individuals identified in a second set of genealogical data that are ancestors to the second human individual, the common ancestor being an individual identified in both the first and second sets; and
 - providing, by the computer, information about the common ancestor to the first human individual.
2. The method of claim 1 where in the genetic data sampled from the first human individual includes values of single nucleotide polymorphisms present within DNA of the first human individual.
3. The method of claim 1 wherein identifying the second human individual genetically related to the first human individual further comprises:
 - identifying shared segments of genotype data between the first human individual and the second human individual, the shared segments exceeding a threshold level.
4. The method of claim 3 wherein the threshold level is exceeded responsive to the first human individual and the second human individual having at least $5 * 10^6$ identical continuous base pairs.

5. The method of claim 1 wherein the first set of genealogical data describes familial relationships of the first human individual received from the first individual and the second set of genealogical data describes familial relationships of the second human individual received from the second individual.

6. The method of claim 5 further comprising:
identifying, by the computer, an additional individual identified in the first set of genealogical data genealogically related to the second human individual, the additional individual not identified in the second set; and
providing, by the computer, information about the additional individual to the second human individual.

7. The method of claim 6 further comprising:
adding, by the computer, indicia of the additional individual to the second set of genealogical data in response to a request from the second individual.

8. The method of claim 1 wherein comparing individuals identified in the first and second sets of genealogical data includes comparing bibliographic data about the individuals.

9. The method of claim 8 wherein the bibliographic data includes places of residence of the individuals.

10. A computer-implemented method for identifying genealogically related individuals, the method comprising:
identifying, by a computer, a plurality of records stored in a genealogical database, each record identifying a unique human individual;

for each of the identified unique human individuals, creating a descendent group, the descendent group including records stored in the genealogical database for each known descendant of the identified unique human individual;

for each descendant group, creating an associated family network, the family network including records of individuals in the descendant group having shared segments of genotype data with at least one other individual in the descendant group, the shared segments exceeding a threshold level.

11. The method of claim 10 further comprising:

receiving genetic data sample from a first human individual;

identifying, by the computer, using the received genetic data, a plurality of human individuals genetically related to the first human individual, each of the plurality of identified individuals belonging to one of the family networks; and

providing to the first human individual indicia of at least one of the identified plurality of human individuals as a suggested genetic relation.

12. The method of claim 11 wherein providing indicia of the identified plurality of human individuals further comprises:

determining according to a confidence score whether a number of human individuals in the identified plurality, each belonging to a same family network, is significant; and

responsive to the number being significant, providing the indicia.

13. The method of claim 12 wherein the confidence score is determined according to a ratio of the number of individuals in the family network genetically related to the first individual to a total number of individuals in the family network.

14. The method of claim 13 wherein the confidence score is further determined according to a total number of individuals in the plurality of identified individuals genetically related to the first human individual.

15. The method of any of claims 1-9 or 11-14 further comprising providing to the first individual indicia of a suggested geographic location associated with an ancestor of the first individual.

16. The method of any of claims 1-9 or 11-14 further comprising providing to the first individual indicia of a suggested surname associated with an ancestor of the first individual.

17. A computer-implemented method for identifying genealogically related individuals, the method comprising:

receiving genetic data sampled from a first human individual;

identifying, by a computer, using the received genetic data, a second human

individual genetically related to the first human individual, thereby indicating a suggested relationship between the first human individual and the second human individual;

identifying a family network associated with the second individual, the family

network including records for individuals having shared segments of genotype data with the second individual, at least some of the shared segments exceeding a threshold level;

determining bibliographic information about at least one individual in the family

network for the second human individual; and

providing the bibliographic information to the first individual.

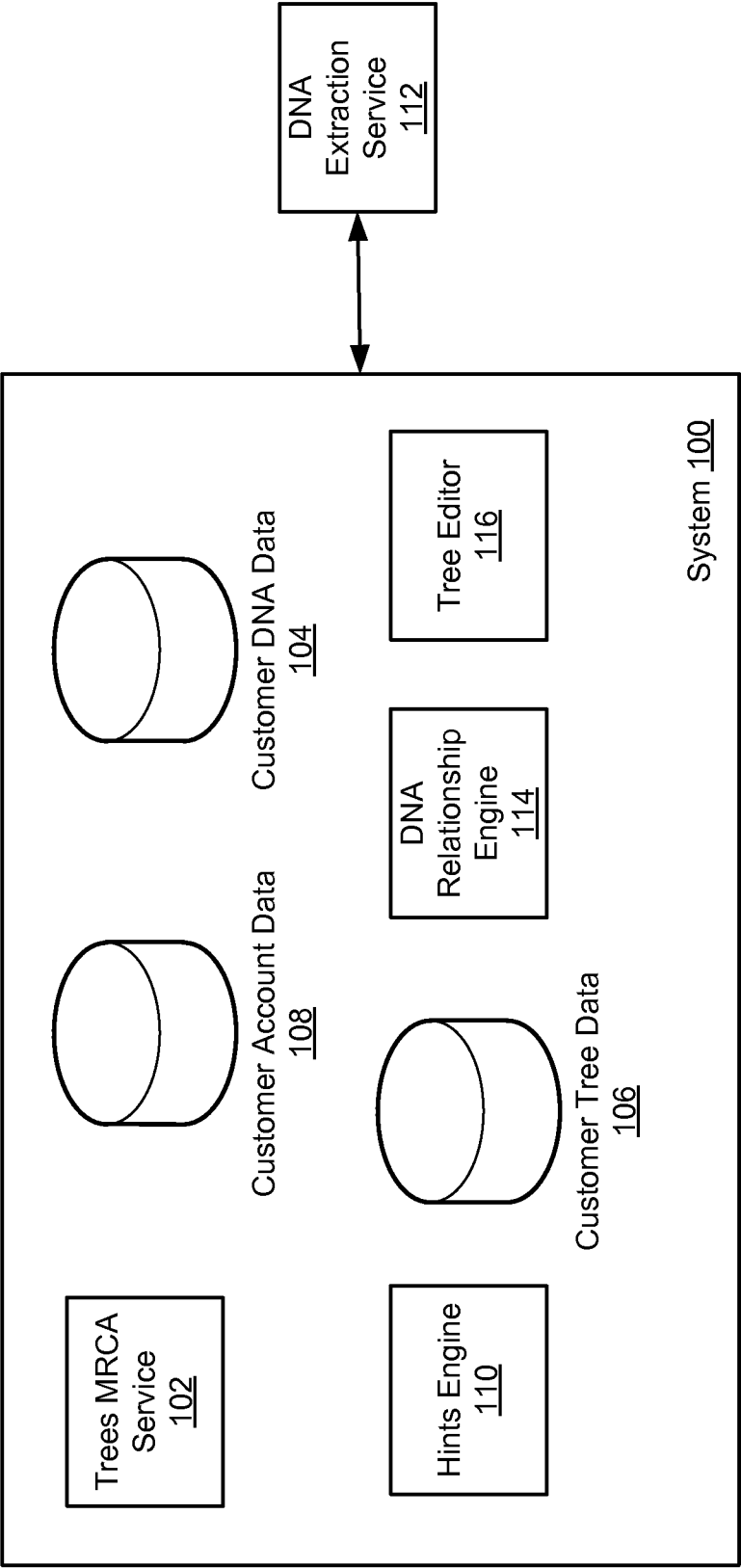


Fig. 1

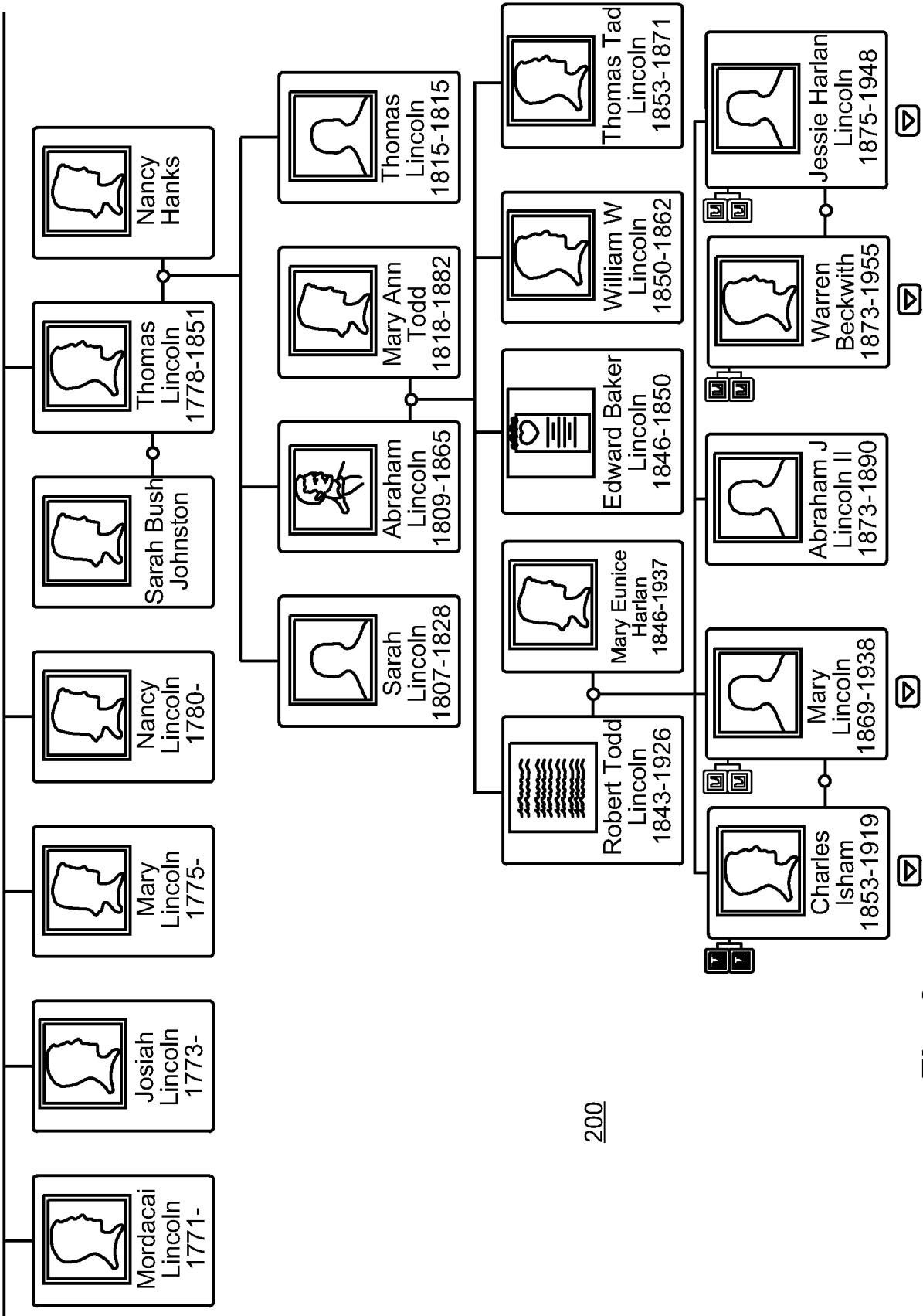
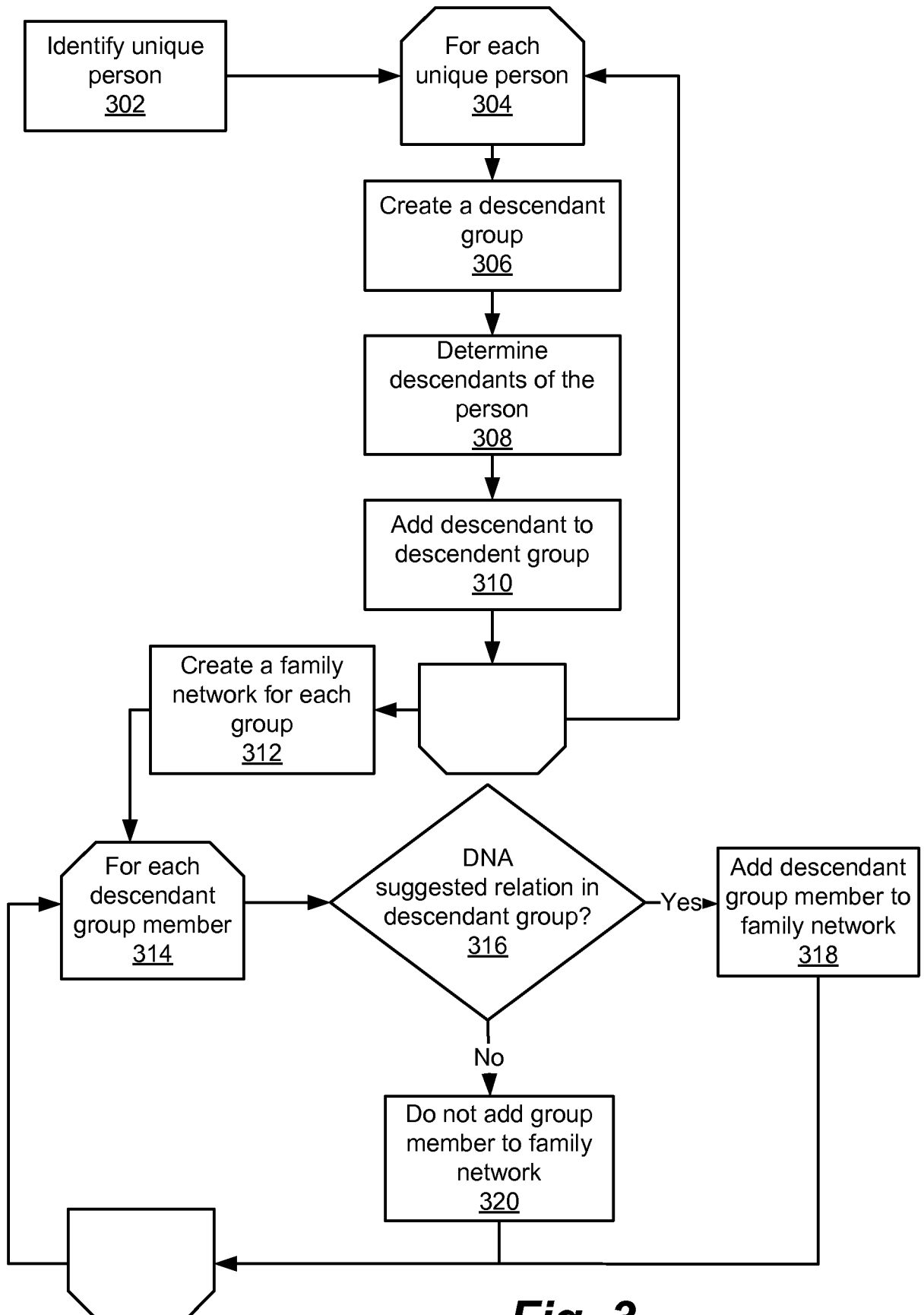
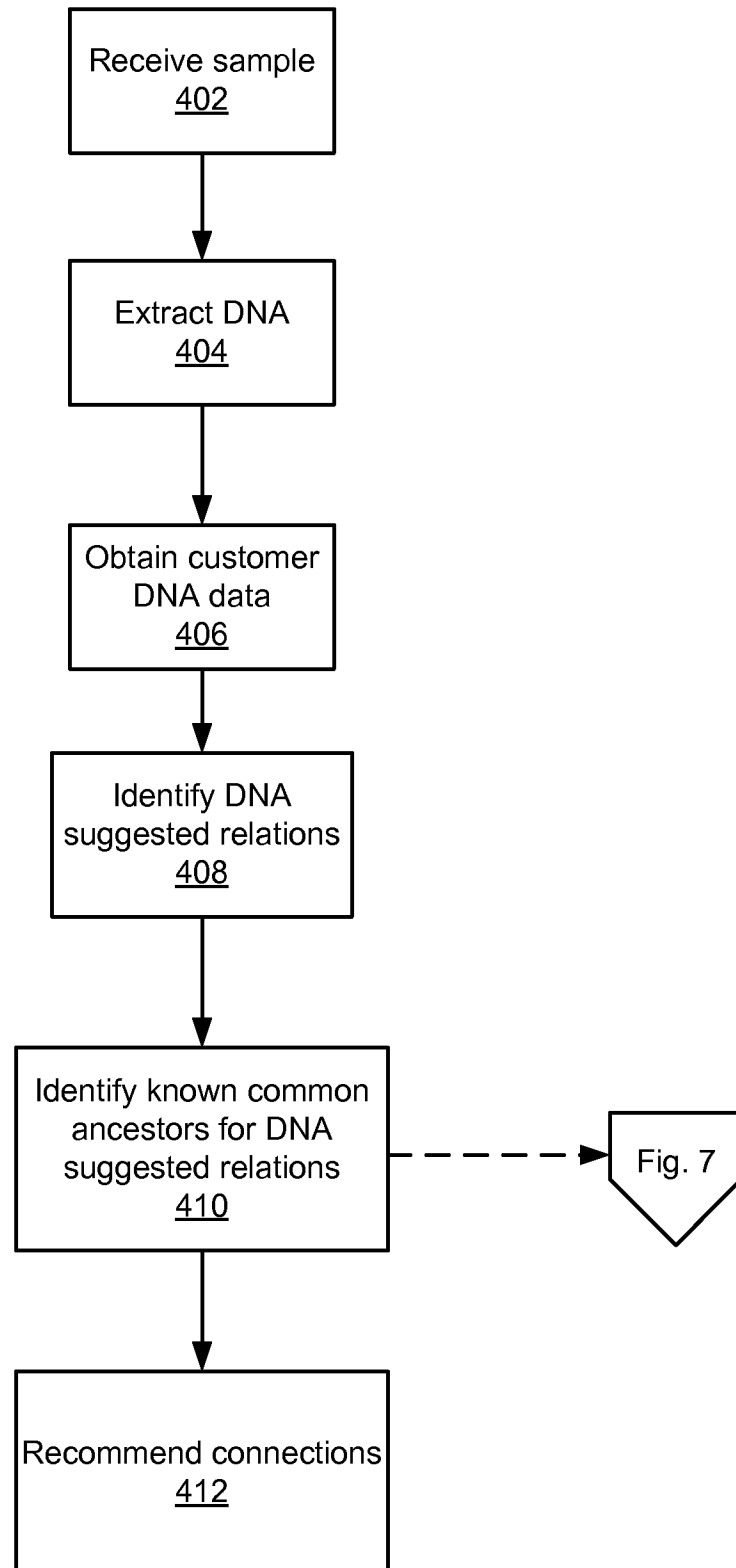
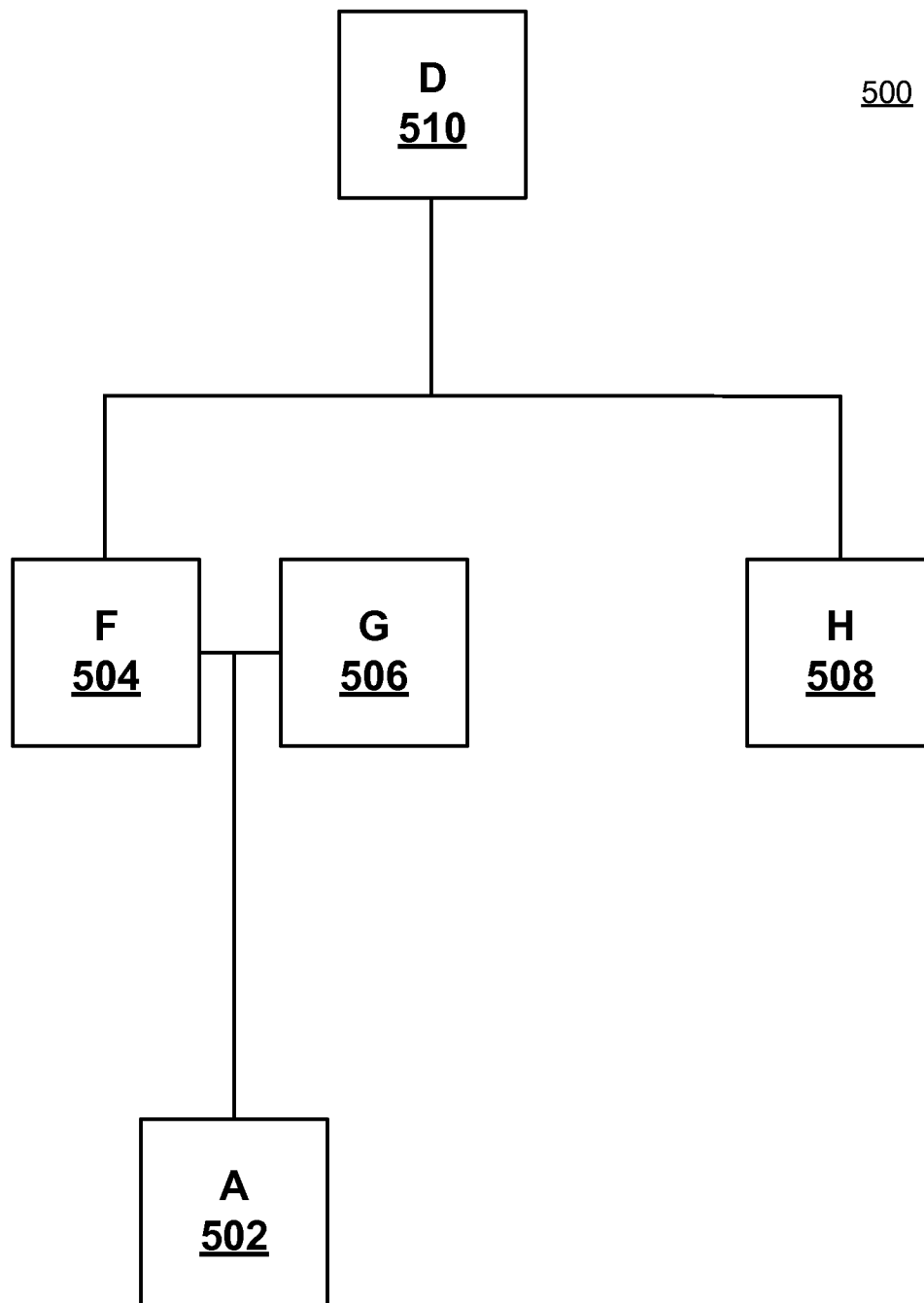
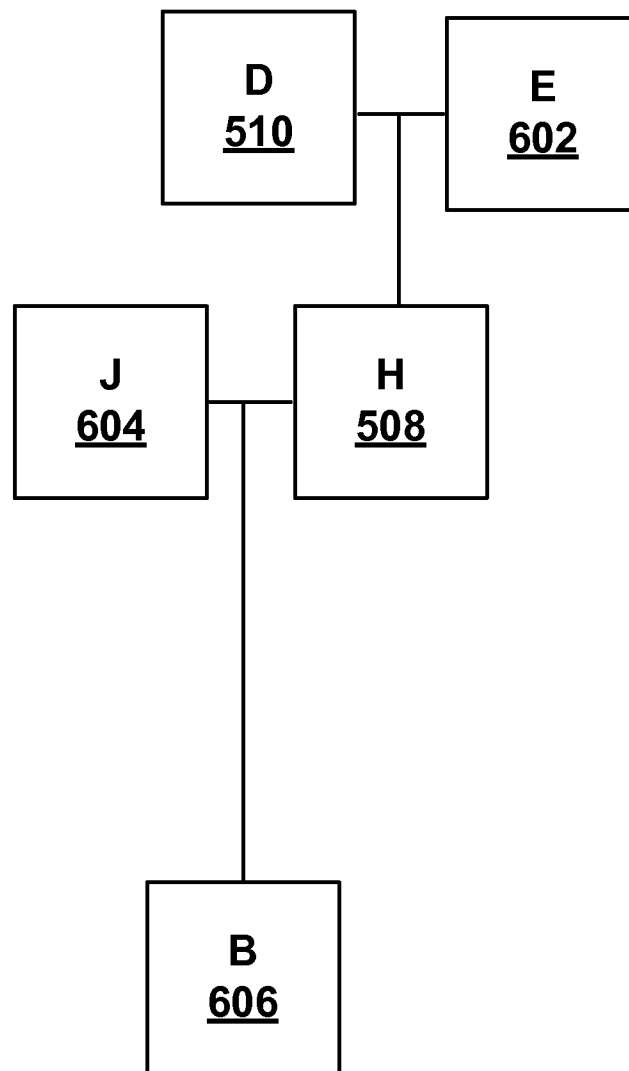


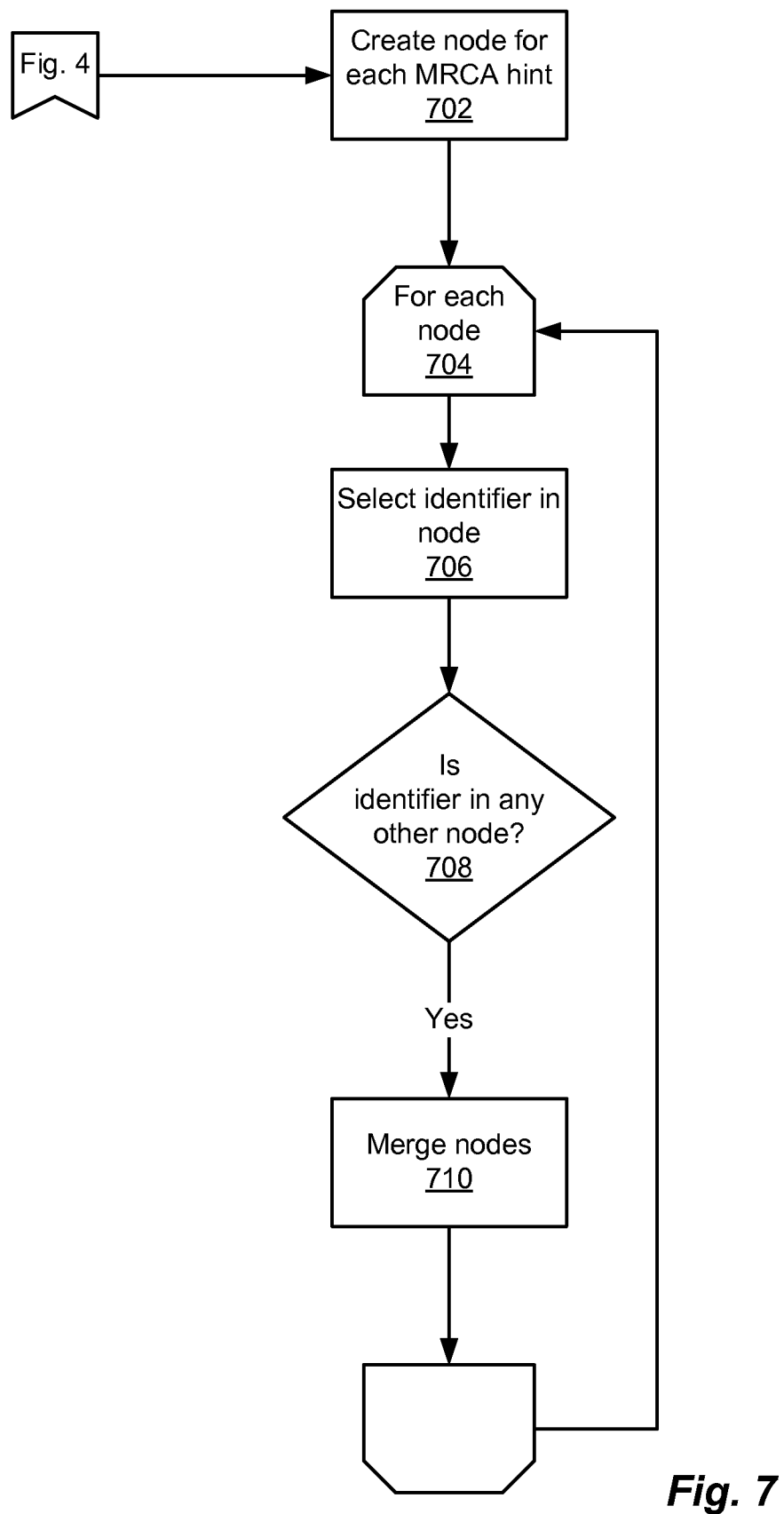
Fig. 2

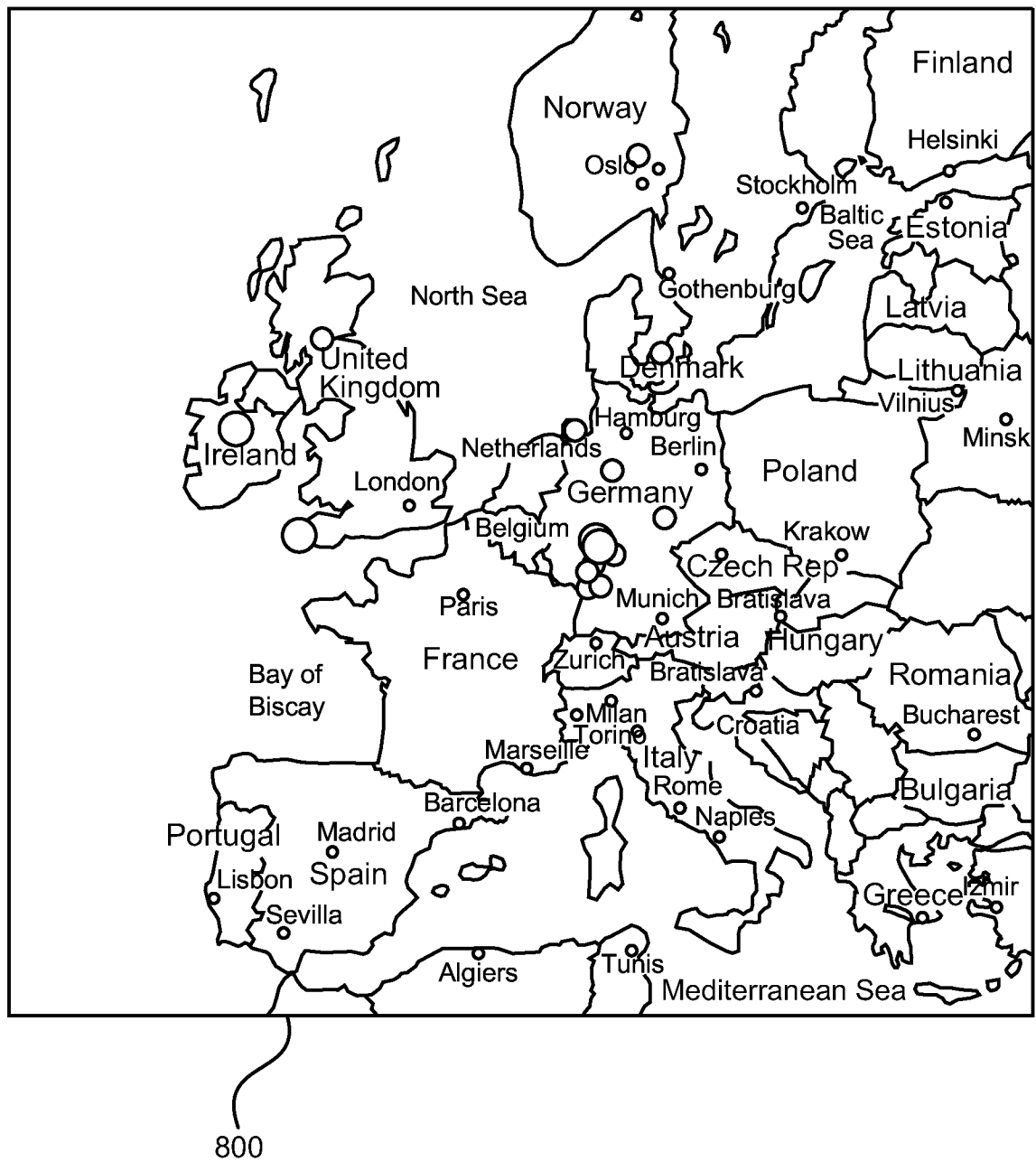
**Fig. 3**

**Fig. 4**

**Fig. 5**

600***Fig. 6***



**Fig. 8**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/030014**A. CLASSIFICATION OF SUBJECT MATTER****G06F 19/10(2011.01)i**

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)

G06F 19/10; G06F 19/00; G06F 17/30; G06F 7/00; G06Q 50/00; G06Q 10/00; G09B 019/00; C12Q 1/68

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: genealogical data, common ancestor, genetic, and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2006-0025929 A1 (CHRIS EGLINGTON) 02 February 2006 See paragraphs [0002]-[0005] and [0038]-[0039]; and claim 1.	1-17
A	US 2002-0032687 A1 (KENT W. HUFF) 14 March 2002 See paragraph [0040] and claim 19.	1-17
A	US 2008-0027656 A1 (LAXMI PRIYA PARIDA) 31 January 2008 See paragraphs [0022] and [0058]-[0072]; and figure 2.	1-17
A	US 2005-0147947 A1 (BENNETT COOKSON JR. et al.) 07 July 2005 See paragraphs [0008] and [0033]-[0049]; and figure 2A.	1-17
A	US 2007-0178500 A1 (LUCAS MARTIN et al.) 02 August 2007 See paragraphs [0010] and [0111]-[0120]; and figure 3.	1-17



Further documents are listed in the continuation of Box C.



See patent family annex.

* 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

10 July 2014 (10.07.2014)

Date of mailing of the international search report

11 July 2014 (11.07.2014)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/030014

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006-0025929 A1	02/02/2006	None	
US 2002-0032687 A1	14/03/2002	US 6760731 B2	06/07/2004
US 2008-0027656 A1	31/01/2008	US 8700334 B2	15/04/2014
US 2005-0147947 A1	07/07/2005	EP 1550958 A2	06/07/2005
		EP 1550958 A3	07/06/2006
		US 2013-268564 A1	10/10/2013
US 2007-0178500 A1	02/08/2007	US 2008-0255768 A1	16/10/2008
		US 8285486 B2	09/10/2012
		WO 2007-084902 A2	26/07/2007
		WO 2007-084902 A3	27/11/2008
		WO 2007-084902 A9	15/01/2009