RZECZPOSPOLITA POLSKA

(12) **TŁUMACZENIE PATENTU EUROPEJSKIEGO**

(19) **PL**

(11) **PL/EP 3831011**

Urząd Patentowy Rzeczypospolitej Polskiej

(96) Data i numer zgłoszenia patentu europejskiego:
12.08.2019 19766302.4

(97) O udzieleniu patentu europejskiego ogłoszono:
01.12.2021 Europejski Biuletyn Patentowy 2021/48
EP 3831011 B1

(13) **T3**

(51) Int.Cl.
H04L 9/08 (2006.01)
H04L 9/32 (2006.01)

(54) Tytuł wynalazku:

Identyfikacja chromosomów

(30) Pierwszeństwo:
10.08.2018 GB 201813130

(43) Zgłoszenie ogłoszono:

09.06.2021 w Europejskim Biuletynie Patentowym nr 2021/23

(45) O złożeniu tłumaczenia patentu ogłoszono:

21.03.2022 Wiadomości Urzędu Patentowego 2022/12

(73) Uprawniony z patentu:

CROALL, Paul Andrew, Bourne End, GB

(72) Twórca(y) wynalazku:

PAUL ANDREW CROALL, Bourne End, GB

(74) Pełnomocnik:
rzecz. pat. Maciej Czarnik
TRILOKA CZARNIK OŻÓG KANCELARIA PATENTOWA I ADWOKACKA SP. P.
ul. K. Ujejskiego 12/7
30-102 Kraków

PL/EP 3831011 T3

Opis

Dziedzina

[0001] Niniejszy wynalazek dotyczy identyfikacji i komunikacji z osobami genetycznie podobnymi. W szczególności niniejszy wynalazek dotyczy sposobu, urządzenia i układu do komunikacji z członkami rodziny użytkownika przy użyciu DNA użytkownika bez upubliczniania profilu DNA.

Tło

[0002] Profil kwasu dezoksyrybonukleinowego (DNA) każdego człowieka może działać jako bardzo specyficzny marker identyfikujący tego konkretnego człowieka. Niektórzy ludzie, na przykład ci blisko spokrewnieni biologicznie, mogą dzielić większą część swojego DNA z takimi członkami rodziny w porównaniu z bardziej oddalonymi członkami rodziny. Dlatego porównując profil DNA dwóch lub więcej osób, można dokładnie określić bliskość spokrewnienia biologicznego między tymi osobami.

[0003] Znalezienie członków rodziny, zwłaszcza zmarłych byłych członków rodziny, może być bardzo trudne, gdy brakuje informacji fizycznych, takich jak przybliżona lokalizacja lub imię.

[0004] Zgłoszenie patentowe US 2015/112884 A1 ujawnia sposoby określania pokrewieństwa między genomami oraz sposoby publicznego udostępniania bezpiecznych szkiców genomów.

[0005] Zgłoszenie patentowe US 2014/289536 A1 ujawnia sposoby, układy i urządzenia do szyfrowania i porównywania danych genomowych. Szyfrowanie danych genomowych umożliwia transmisję, przechowywanie i wykorzystywanie danych genomowych na bezpiecznym nośniku.

Podsumowanie Wynalazku

[0006] Aspekty i/lub przykłady wykonania mają na celu zapewnienie sposobu, urządzenia i układu do komunikacji z członkami rodziny użytkownika przy użyciu DNA użytkownika bez upubliczniania profilu DNA, jak określono w załączonych zastrzeżeniach.

[0007] Według pierwszego aspektu, zapewniony jest realizowany komputerowo sposób lokalizowania jednego lub większej liczby członków sieci rodzinnej, obejmujący etapy: generowanie jednego lub większej liczby kluczy szyfrowania pochodzących z pierwszej sekwencji genomowej; szyfrowanie wiadomości przy użyciu jednego lub każdego klucza szyfrowania w celu utworzenia zaszyfrowanej wiadomości; wysyłanie zaszyfrowanej wiadomości do jednego lub większej liczby zdalnych urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie zdalnych urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania pochodzących z drugiej sekwencji genomowej; oraz odbieranie potwierdzenia dotyczącego tego, czy odszyfrowanie zaszyfrowanej wiadomości zakończyło się powodzeniem przez którekolwiek z jednego lub więcej zdalnych urządzeń, przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej.

[0008] Opcjonalnie, ujawniony tu sposób obejmuje ponadto etapy: otrzymywania danych wejściowych zawierających pierwszą sekwencję genomową; oraz generowanie jednego lub większej liczby kluczy szyfrowania na podstawie danych wejściowych. Opcjonalnie etap odszyfrowania zaszyfrowanej wiadomości na jednym lub większej liczbie zdalnych urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania pochodzących z jednej lub większej liczby dalszych sekwencji genomowych.

[0009] Ujawniony tutaj układ, określany również jako układ identyfikacji chromosomów, zapewnia możliwość komunikowania się z potencjalnie utraconymi członkami rodziny przy użyciu DNA nadawcy bez upubliczniania żadnego DNA. Nadawca (lub „użytkownik") może obejmować osobę, która wprowadza swój DNA do układu identyfikacji chromosomów w celu wykorzystania układu dostarczonego w niniejszym dokumencie. Na przykład zaginiony członek rodziny może zostać zidentyfikowany i może dojść do spotkania. Ponieważ dane genomowe można uznać za prywatne, ważne jest, aby bardzo ostrożnie wybierać osoby, które mogą uzyskać do nich dostęp. W tym celu do utworzenia adresu genetycznego można wykorzystać zsekwencjonowany genom zawierający lub pochodzący z DNA użytkownika. Użytkownik może przygotować dla nich układ identyfikacji chromosomów w celu zaszyfrowania wiadomości. Metryka używana do szyfrowania jest zakodowana w genomie lub w jednej lub kilku właściwościach genomu. W szczególności, w jednym przykładzie wykonania właściwości genomu mogą obejmować właściwości mitochondriów oraz chromosomów X i Y. Wiadomość nie jest samą sekwencją, ponieważ ma być utrzymywana w tajemnicy, ale jest to niektóre inne dane (np. wiadomość) zaszyfrowane przy użyciu klucza szyfrującego pochodzącego z sekwencji genomowej.

[0010] Opcjonalnie, jeden lub więcej kluczy szyfrowania zawiera: pierwszy klucz szyfrowania dla algorytmu szyfrowania dosłownego i drugi klucz szyfrowania dla unikalnego algorytmu szyfrowania. Opcjonalnie pierwszy klucz szyfrowania jest używany w odniesieniu do: jednego lub większej liczby mitochondriów; pierwszy chromosom X; oraz drugi chromosom X lub chromosom Y, aby utworzyć trzy dosłownie generowane algorytmicznie szyfry w odniesieniu do pierwszej sekwencji genomowej. Opcjonalnie drugi klucz szyfrowania jest używany do zaszyfrowania wiadomości przy użyciu kombinacji chromosomów od 1 do 22 zawartych w pierwszej sekwencji genomowej. Opcjonalnie etap wysyłania zaszyfrowanej wiadomości jest wykonywany przy użyciu jednego lub większej liczby danych wyjściowych pochodzących z pierwszej sekwencji genomowej.

[0011] Dosłowna kryptografia stosowana jako część algorytmu dosłownego szyfrowania wykorzystuje fakt, że te trzy chromosomy mogą być przekazywane z rodzica na potomstwo, z pewnymi zastrzeżeniami. Chromosomy od 1 do 22 razem, z rzadkimi wyjątkami, są konwencjonalnie akceptowane jako unikalne dla każdego osobnika. Szyfrowanie można zatem przeprowadzić przy użyciu chromosomów, które wydają się być unikalne dla każdego osobnika, ale biorąc pod uwagę fakt, że części wyżej wymienionych chromosomów mogą być dzielone z krewnymi.

[0012] Opcjonalnie udane odszyfrowanie zaszyfrowanej wiadomości wskazuje na z góry określony poziom spokrewnienia. Opcjonalnie pierwsza i/lub jedna lub więcej dalszych sekwencji genomowych zawiera co najmniej część sekwencji genomu. Opcjonalnie pierwsza i/lub druga sekwencja genomowa zawiera co najmniej część sekwencji genomu

[0013] Biorąc pod uwagę, że części wyżej wymienionych chromosomów mogą być współdzielone z krewnymi, udane odszyfrowanie zaszyfrowanej wiadomości musi oznaczać, że dane z DNA użytkownika deszyfrującego wiadomość zawierają co najmniej część wspólną z DNA użytkownika szyfrującego wiadomość.

[0014] Opcjonalnie sekwencja genomowa zawiera cztery zasady zawierające jedną lub więcej spośród: guaniny (G), cytozyny (C), adeniny (A), tyminy (T) i/lub uracylu (U).

[0015] Te zasady są używane jako części składowe konwencjonalnej sekwencji genomowej, ponieważ są to zasady używane do tworzenia DNA.

[0016] Opcjonalnie cztery zasady są mapowane na sekwencję binarną. Opcjonalnie sekwencja binarna zawiera dwie formy: skompresowaną i/lub nieskompresowaną. Opcjonalnie wejście zawiera liczbę jednej lub więcej z czterech zasad.

[0017] Każdy chromosom może być przechowywany w postaci binarnej, 4 bity na cyfrę o podstawie 4, a operacje kryptograficzne i bitowe mogą być na nich wykonywane w celu użycia ich jako kluczy. Takie użycie może obejmować wciskanie ich do postaci, która może być postrzegana jako 2 bity na 4 cyfrę bazową. Ta struktura pozwala uniknąć nadmiernego stronicowania nieistotnej ilości danych, biorąc pod uwagę cały ludzki genom, zmniejszając w ten sposób zużycie pamięci RAM w czasie wykonywania sekwencjonowanych chromosomów. Sposób ten można określić jako produkt DNA lub ProDNA.

[0018] Opcjonalnie odszyfrowanie zaszyfrowanej wiadomości jest pomyślne tylko wtedy, gdy z góry określona proporcja pierwszej sekwencji genomowej odpowiada jednej lub większej liczbie dalszych sekwencji genomowych. Opcjonalnie z góry określona proporcja odzwierciedla poziom pokrewieństwa między właścicielem pierwszej sekwencji genomowej a właścicielem jednej lub większej liczby dalszych sekwencji genomowych. Opcjonalnie odszyfrowanie zaszyfrowanej wiadomości jest pomyślne tylko wtedy, gdy z góry określona proporcja pierwszej sekwencji genomowej odpowiada drugiej sekwencji genomowej. Opcjonalnie z góry określona proporcja odzwierciedla poziom pokrewieństwa między właścicielem pierwszej sekwencji genomowej a właścicielem drugiej sekwencji genomowej.

[0019] Jeśli dwie osoby mają więcej niż wcześniej określoną proporcję ich sekwencji genomowych, prawdopodobnie wskazuje to na biologiczną bliskość między tymi dwoma osobami. Korzystne może być skontaktowanie się tylko z krewnymi, w przeciwnym razie zostanie wysłana duża liczba fałszywych wiadomości i zmniejszona zostanie dokładność odnalezienia krewnych.

[0020] Teraz zostaną wprowadzone dwie koncepcje, aby ułatwić zrozumienie poniższego opisu i przykładów wykonania opisanych później. Te dwa pojęcia to „głos” i „głośność”. W tym kontekście „głos” nie dotyczy ilości DNA, dotyczy preferencji w odniesieniu do docelowych odbiorców danej komunikacji. Skalę, na której mierzy się głośność dla celów tego opisu, można nazwać „głośnością”. Im głośniejszy dźwięk, tym dalej można go usłyszeć, biorąc pod uwagę dystans genetyczny. Ustawiając głośność w taki sposób, w jaki można ustawić głośność w systemie muzycznym na głośną lub cichą lub gdziekolwiek pomiędzy, określa się, kto może usłyszeć wiadomość (tj. docelowi odbiorcy danej komunikacji). Głośność określa, jak daleko użytkownik może być genetycznie od nadawcy (tj. dystans genetyczny) i nadal „słyszeć” jego wiadomość. Podczas gdy w układzie muzycznym głośność zazwyczaj dotyczy jednej skali, tutaj może występować wiele skal lub właściwości, które są regulowane w sposób zrównoważony, zgodnie z ustawieniem głośności. Ustawienie objętości można uznać za prostszy sposób uzgodnienia przez nieznajomych złożonego zestawu szczegółów kalibracji. Zamiast brać pod uwagę takie zmienne, mogą po prostu ustawić głośność.

[0021] W związku z ustawieniem głośności użytkownik może poczynić następujące dwie uwagi:

1. Podczas wysyłania wiadomości im „głośniej” ustawiona jest głośność, tym do większej liczby osób dotrze ich wiadomość, ponieważ nawet dalsi krewni są w stanie skutecznie odszyfrować wiadomość; i odwrotnie, i

2. Podczas odbierania wiadomości im głośniejsza jest ustawiona głośność, tym więcej wiadomości zostanie odebranych od jeszcze bardziej odległych krewnych i odwrotnie.

[0022] W zasadzie bliźniaki będą mieli identyczne DNA. Nawet pojedyncza różnica par zasad, być może wprowadzona przez błąd sekwencjonowania, może wystarczyć do ich rozróżnienia, ale taki błąd nie może wystąpić. W takim przykładowym scenariuszu, każde identyczne rodzeństwo może zastosować ujawniony tutaj sposób, aby znaleźć tę samą osobę. Pomimo podobieństwa genetycznego, jedyną różnicą może być to, że ich genom nie zostanie zsekwencjonowany dokładnie w tym samym momencie. Dlatego do rozróżnienia każdego rodzeństwa można użyć wystarczająco dokładnego znacznika czasu z każdego pliku sekwencji. Ten mechanizm rozróżniania identycznego rodzeństwa może być rozwiązaniem nieidealnym w porównaniu z konwencjonalnym przypadkiem bez identycznego rodzeństwa. Jednak nadal może być ważne zapewnienie takiego wyróżnienia. Należy zauważyć, że ten konkretny sposób rozróżniania identycznego rodzeństwa może reprezentować jedną z kilku opcji, które można zastosować, a inne środki mogą być oczywiste dla specjalisty w dziedzinie.

[0023] Opcjonalnie etap wysyłania zaszyfrowanej wiadomości obejmuje utworzenie układu liczb całkowitych trzy na cztery oraz unikalnego indywidualnego zaszyfrowanego identyfikatora liczb całkowitych. Opcjonalnie unikalna pojedyncza zaszyfrowana liczba całkowita jest skrótem sekwencji

całego genomu. Opcjonalnie, unikalna indywidualna zaszyfrowana liczba całkowita zawiera ponadto nałożenie znacznika czasu daty sekwencjonowania.

[0024] Informacje związane z DNA, takie jak sekwencja genomowa, mogą być bardzo osobistymi informacjami, których użytkownik może nie chcieć udostępniać na szeroką skalę. Dlatego jest korzystne, jeśli dokładny charakter takich informacji może pozostać zaszyfrowany lub w inny sposób zakodowany, nawet w przypadku użycia do identyfikacji krewnych. Użytkownik może nie w pełni ufać wspomnianym krewnym i dlatego może chcieć zachować takie dane osobowe tylko dla siebie.

[0025] Opcjonalnie, etap szyfrowania wiadomości obejmuje użycie binarnego dużego obiektu słów o pierwszej wielkości (BoPSW).

[0026] Jak szczegółowo opisano w niniejszym dokumencie, BoPSW może zapewnić bezpieczne i wydajne obliczeniowo środki szyfrowania w odniesieniu do danych wykorzystywanych w układzie identyfikacji chromosomów.

[0027] Opcjonalnie jeden lub każdy z węzłów rozproszonej sieci węzłów jest zdolny do wysłania zaszyfrowanej wiadomości tylko za pierwszym razem, gdy zaszyfrowana wiadomość zostanie odebrana przez ten lub każdy węzeł.

[0028] W jednym przykładzie wykonania węzeł może przekazywać wiadomość tylko za pierwszym razem, gdy ją odbierze, i nie pozwoli na odebranie po raz drugi, chociaż jeśli otrzyma nowe tłumaczenie wiadomości, może również zostać przekazana. Po zgrupowaniu wszystkich tych wiadomości do przekazywania węzła zawiera szczegóły potencjalnej powiększonej rodziny. Jednak wielkość rodziny będzie ograniczona przez jeden lub więcej szczegółów kalibracji głośności przekazywanej wiadomości. Powyższy układ może być ładowany przez, zanim odbierze jakiekolwiek wiadomości, każdy węzeł ma wstępnie skonfigurowaną bramę lub korzystnie dwa inne węzły. Jeżeli właściciel węzła chce wysłać wiadomość, zanim węzeł kiedykolwiek odbierze wiadomość, wiadomość zostanie przekazana do bramy, ponieważ w co najmniej jednym przykładzie wykonania nie ma innego miejsca do jej przekazania. Gdy węzeł wyśle wiadomość, jej adres staje się znany sieci, jak opisano powyżej, i od tego momentu może odbierać ukierunkowane wiadomości.

[0029] Opcjonalnie, każdy węzeł sieci rozproszonej ma skojarzonego użytkownika.

[0030] Użytkownik może być połączony z określonym węzłem w celu identyfikacji tego użytkownika.

[0031] Opcjonalnie, ujawniony tu układ obejmuje ponadto etap: wyprowadzania miary odległości genetycznej między pierwszą sekwencją genomową a jedną lub większą liczbą dalszych sekwencji genomowych. Opcjonalnie, ujawniony tu układ obejmuje ponadto etap: wyprowadzania miary odległości genetycznej między pierwszą sekwencją genomową a drugą sekwencją genomową. Opcjonalnie miara odległości genetycznej jest określana za pomocą adresu DNA.

[0032] Użytkownik może chcieć korespondować tylko z innymi użytkownikami w określonym zakresie biologicznym. Dlatego może być korzystne zapewnienie takiego środka, który będzie zrozumiały dla użytkownika.

[0033] Opcjonalnie zaszyfrowana wiadomość jest przechowywana przez pewien czas na zdalnym serwerze. Opcjonalnie odszyfrowanie zaszyfrowanej wiadomości po etapie wysłania zaszyfrowanej wiadomości następuje z opóźnieniem czasowym.

[0034] Użytkownicy układu identyfikacji chromosomów mogą fizycznie znajdować się na dużych odległościach, a zatem być częścią różnych sieci i/lub serwerów. Jednak mogą nadal chcieć używać aranżacji wraz z ich fizycznie odległymi relacjami.

[0035] Według kolejnego aspektu, zapewniono urządzenie do rodzinnej lokalizacji sieci, zawierające: procesor zdolny do: generowania jednego lub większej liczby kluczy szyfrowania pochodzących z pierwszej sekwencji genomowej; zaszyfrować wiadomość przy użyciu jednego lub każdego klucza szyfrowania w celu utworzenia zaszyfrowanej wiadomości; wysłać zaszyfrowaną wiadomość do jednego lub większej liczby urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania pochodzących z drugiej sekwencji genomowej; i otrzymać potwierdzenie dotyczące tego, czy odszyfrowanie zaszyfrowanej wiadomości powiodło się przez jedno lub więcej zdalnych urządzeń, przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej.

[0036] Według dalszego aspektu zapewniony jest układ do lokalizowania jednego lub większej liczby członków sieci rodzinnej, obejmujący: procesor zdolny do: generowania jednego lub większej liczby kluczy szyfrowania pochodzących z pierwszej sekwencji genomowej; zaszyfrowania wiadomości przy użyciu jednego lub każdego klucza szyfrowania w celu utworzenia zaszyfrowanej wiadomości; wysłania zaszyfrowanej wiadomości do jednego lub większej liczby urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania pochodzących z drugiej sekwencji genomowej; i otrzymania potwierdzenia dotyczącego tego, czy odszyfrowanie zaszyfrowanej wiadomości powiodło się przez jedno lub więcej zdalnych urządzeń, przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej.

[0037] Urządzenie i układ umożliwiają wykonanie sposobu identyfikacji chromosomów i związanej z nią transmisji wiadomości.

Krótki Opis Figur Rysunku

[0038] Przykłady wykonania zostaną teraz opisane jedynie tytułem przykładu i w odniesieniu do załączonych figur rysunku mających podobne numery odniesienia, na których:

Fig. 1 przedstawia ogólny schematyczny przegląd układu, jak tu ujawniono;

Fig. 2 przedstawia przykład obliczenia odległości między członkami rodziny;

Fig. 3 przedstawia przykładową reprezentację części składowych DNA w postaci binarnej;

Fig. 4 przedstawia przykładowe obliczenie produktu DNA;

Fig. 5 przedstawia operację wieloskokową; oraz

Fig. 6a, 6b i 6c przedstawiają przykładową postać wykonania ujawnionego tutaj układu.

Szczegółowy Opis

[0039] W odniesieniu do fig. 1, zostanie teraz opisany pierwszy przykład wykonania. Zsekwencjonowany genom 106 zawierający DNA użytkownika jest wykorzystywany do tworzenia adresu genetycznego 105 do celów routingu. Następnie użytkownik wykorzystuje układ identyfikacji chromosomów do zaszyfrowania wiadomości 100 w zaszyfrowaną wiadomość 104, 140', która może następnie zostać wysłana do potencjalnie utraconych członków rodziny, którzy mogą nawet nie znać nazwiska odbiorcy, przybliżonego miejsca pobytu ani żadnych innych istotnych informacji o śledzeniu. Klucz 140, 150 używany do szyfrowania pochodzi 108 z genomu lub z jednej lub więcej właściwości genomu. W szczególności, w jednym przykładzie wykonania właściwości genomu obejmują właściwości mitochondriów oraz chromosomów X i Y. Wielu użytkowników będzie miało identyczne kopie pasujących mitochondriów i/lub chromosomów X i/lub Y ich bliskich krewnych.

[0040] Człowiek ma trzy kluczowe korzenie genetyczne. W przypadku mężczyzn każdy z tych korzeni może być uważany za zasadzony w podłożu (termin, który może być używany zamiennie z „domeną" lub „własnością") własnego typu, zawierającym tylko ten rodzaj korzenia. Kobiety mają jeden z korzeni dzielonych z jednym z męskich korzeni, dwa z korzeni dzielonych z innym męskim korzeniem, ale nie dotykają trzeciego podłoża. W tym przykładzie korzenie reprezentują chromosom mitochondriów 115, chromosom X (lub chromosomy) 120 i chromosom Y 125.

[0041] Każda osoba ma swoje trzy korzenie w trzech określonych miejscach na danym terenie. W każdym przypadku lokalizację można zidentyfikować za pomocą współrzędnych czterowymiarowych. Współrzędną można zdefiniować jako ΣG, ΣC, ΣA, ΣT, gdzie Σ służy do wskazania numeru tej konkretnej zasady w danym chromosomie, gdzie G, C, A i T dotyczą odpowiednio guaniny, cytozyny, adeniny i tyminy. To obliczenie można podsumować w tabeli 105.

[0042] Można założyć, że określone trzy zestawy współrzędnych nie są udostępniane większości ludzi. Tacy ludzie, z który nie ma wspólnych współrzędnych, zgodnie z konkretną definicją bliskiego krewnego, nie są bliskimi biologicznymi krewnymi użytkownika, od którego pobrano DNA w celu wygenerowania tych konkretnych trzech zestawów współrzędnych. Nieliczni, którzy dzielą z użytkownikiem co najmniej jedną ze swoich lokalizacji genetycznych, mają większą szansę na bycie

bliskim krewnym biologicznym. Jeśli dana osoba znajduje się tylko lub dwie mutacje od lokalizacji genetycznej użytkownika, istnieje znacznie większe prawdopodobieństwo znalezienia bliskiego biologicznego krewnego.

[0043] Różne komunikaty 100' i 100" (lub tyle, ile jest to wymagane) mogą być nadawane przy użyciu tego układu, ale z różnymi ustawieniami, aby umożliwić na przykład inny zamierzony odsłuch. W konkretnym przykładzie wiadomość 101' może być przeznaczona tylko dla bardzo bliskich członków rodziny, a ustawienia szyfrowania tej wiadomości 101' wymagają odszyfrowania wiadomości tylko przez osoby o bardzo podobnych sekwencjach genetycznych do nadawcy, podczas gdy wiadomość 101" może być przeznaczona dla bardziej odległych odbiorców relacji genetycznych jako nadawcy.

[0044] Eliminując większość ludzi, którzy nie są genetycznie blisko użytkownika, pozostaje znacznie mniejsza pula ludzi, którzy z większym prawdopodobieństwem są blisko spokrewnieni genetycznie, a zatem tworzą miarę dystansu rodzinnego 110. Korzystając z powyższego układu współrzędnych, można zapewnić mniej kosztowny obliczeniowo sposób łączenia tych, którzy chcą być połączeni. Przykład sposobu pomiaru bliskości genetycznej przedstawiono na fig. 2.

[0045] Docenia się, że istnieje chęć zachowania prywatności DNA. DNA może reprezentować bardzo osobiste informacje, których wielu użytkowników może nie chcieć udostępniać. Dlatego też, gdy zostanie znaleziony potencjalny biologiczny krewny, wiadomości mogą być wymieniane i szyfrowane przy użyciu każdego z ich genomów jako części odpowiednich kluczy. Tylko pasujący genom umożliwi odszyfrowanie, pomagając w ten sposób wyeliminować wszelkie fałszywie pozytywne wyniki. Ten sposób stanowi alternatywę dla dostarczania nieznajomym, którzy mogą okazać się niepowiązanymi, danych osobowych, takich jak sekwencja genomowa.

[0046] Układ identyfikacji chromosomów umożliwia wywnioskowanie z genomu osobnika adresu DNA, który może być częściowo przedstawiony w tabeli 105:

	G	C	A	T
M	Σ	Σ	Σ	Σ
X	Σ	Σ	Σ	Σ
X/Y	Σ	Σ	Σ	Σ

Σ oznacza, ile tej zasady znajduje się w chromosomie po pomnożeniu przez 12.

[0047] Czasami nie ma pewności, która zasada jest reprezentowana w określonej sekwencji genomu. Jeśli podstawa jest znana z określonym z góry stopniem pewności, to każda liczba każdej znanej podstawy liczy się jako 12. Jeśli podstawa nie jest znana z określonym z góry stopniem pewności i może być dowolną z 4 zasad, to każda z nich liczy się jako 3. Jeśli podstawa może być tylko jedną z 3

określonych zasad, wtedy każda z tych 3 liczy się jako 4, a pozostała wynosi zero. Jeśli nieznana zasada może być tylko jedną z określonych 2 zasad, to każda z tych 2 liczy się jako 6, a pozostałe 2 są równe zeru.

[0048] Dla celów uproszczonych obliczeń uważa się, że w ludzkim mitochondrialnym DNA zazwyczaj występuje 16 569 par zasad, około 58 000 000 par zasad w chromosomie Y i około 155 000 000 par zasad w chromosomie X. Każda podana podstawa musi mieć formę G, C, A lub T. Jakiekolwiek dwie osoby wybrane losowo z populacji prawdopodobnie będą miały różną liczbę lub ilość danej pary zasad w każdym z ich mitochondriów, chromosomów X i Y. Mnożenie możliwych kombinacji ilości par zasad daje wynik o wiele rzędów wielkości większy niż aktualna populacja ludzka. Wytwarzany jest rozkład krzywej dzwonowej, który można korzystnie stosować. Zamiast tego, że każda osoba ma swoje własne, unikalne miejsce w dystrybucji, blisko spokrewnieni ludzie będą bliżej siebie na krzywej dzwonowej i tak blisko spokrewnieni ludzie będą mieli to samo miejsce w wygenerowanej dystrybucji. Dlatego wyniki mogą być ustawione w uporządkowanej kolejce, a bliscy krewni danej osoby mogą być uważani za znajdujących się w tej samej części tej kolejki.

[0049] Ustanowiono dwanaście oddzielnych takich kolejek:

1) Liczba „G” w chromosomie mitochondrialnym;
2) Liczba „C” w chromosomie mitochondrialnym;
3) Liczba „A” w chromosomie mitochondrialnym;
4) Liczba „T” w chromosomie mitochondrialnym;
5) Liczba „G” w chromosomie X;
6) Liczba „C” w chromosomie X;
7) Liczba „A” w chromosomie X;
8) Liczba „T” w chromosomie X;
9) Liczba „G” w chromosomie Y;
10) Liczba „C” w chromosomie Y;
11) Liczba „A” w chromosomie Y; i
12) Liczba „T” w chromosomie Y.

[0050] Podczas gdy biologiczne kobiety mają dwa chromosomy X, mężczyźni mają jeden chromosom X i jeden Y. Powoduje to, że kobiety mają dwa miejsca w kolejkach X, ale w ogóle nie występują w kolejce Y. Bliski krewny będzie pasował w co najmniej czterech kolejkach, ponieważ będzie miał wspólny chromosom.

[0051] Każda kolejka może zawierać ponad dziesięć tysięcy miejsc. Biorąc pod uwagę osobę, w dowolnej kolejce, bardzo wysoki odsetek osób prawdopodobnie znajdzie się w innym miejscu. Większość kolejek musi mieć o rząd wielkości więcej szczelin, więc proporcja ta staje się jeszcze wyższa. Oznacza to, że dana osoba może dorównać maksymalnie 0,01% populacji ludzkiej. Liczbę tę można obecnie szacować na około 760 000 osób.

[0052] Patrząc tylko na chromosom Y, kolejki te będą o rzędy wielkości większe niż 10 000. Chromosom Y jest o trzy rzędy wielkości dłuższy niż mitochondria. Te trzy rzędy wielkości mogą być wykorzystane do zmniejszenia powyższej liczby 760 000 osób na całym świecie do 760 osób. Chromosom X jest nawet dłuższy niż chromosom Y. Jednak 760 osób to wciąż większa liczba nieznanych osób, niż potencjalny użytkownik może chcieć poinformować o swojej osobistej i potencjalnie bardzo prywatnej sekwencji DNA.

[0053] Do realizacji układu ujawnionego w niniejszym dokumencie można zastosować architekturę komputerową w postaci rzadkiego układu zaimplementowanego przy użyciu drzewa. Daje to korzyści wynikającą z układu oraz drzewa. Drzewiasta implementacja rzadkiego układu może być używana w połączeniu z adresami DNA, aby skojarzyć ze sobą najbardziej spokrewnione osoby. ProDNA implementuje binarną reprezentację chromosomu, z możliwością wykonywania operacji na ProDNA lub między wieloma ProDNA, renderując nowe ProDNA bez potrzeby zużywania dodatkowej pamięci poza początkowym obciążeniem podstawowym i częścią metadanych, która może być stosunkowo mała w porównaniu z początkowym obciążeniem podstawowym.

[0054] Takie operacje są przydatne w kryptografii. Operacje mogą obejmować jedną lub więcej z: XOR, składania i skompresowania, jak przedstawiono w etapie 160. Klasycznie, w czasie ładowania, ProDNA reprezentuje szesnastkową postać sekwencji.

[0055] „PressedProDNA" jest wynikiem naciśnięcia ProDNA. Dosłowna kryptografia, jak tu opisano, może wymagać sekwencjonowania o wyższej jakości, a zatem korzyści z pierwszego naciskania chromosomu. W przeciwieństwie do tego, unikalna kryptografia może być używana wraz z sekwencją o niższej jakości z mniejszymi trudnościami, zwłaszcza gdy ProDNA nie ma wyciśniętych pytań.

[0056] Sposób sekwencjonowania może powodować pewną niepewność co do tego, co dokładnie zostało odczytane. Niesprecyzowane ProDNA mogą stanowić niepewność. Taka niepewność może sprowadzać się do pytania, czy dana para zasad to faktycznie np. G czy T. Wiadomo, że to nie A czy C, ale pozostaje pytanie, czy to G czy T. Przekształcenie nieskompresowanego ProDNA w skompresowane ProDNA wymaga podjęcia decyzji dotyczącej tego niepewnego pytania. W takich okolicznościach decyzja może być oparta na prawdopodobieństwie. Jednak dokładniejszą opcją może być użycie mechanizmu sekwencjonowania wyższej jakości, eliminującego potrzebę korzystania z

prawdopodobieństw. Sekwencje wysokiej jakości można uzyskać powtarzając sposób sekwencjonowania, aż pojawi się klarowność.

[0057] „ProDNAFold" jest wynikiem złożenia ProDNA bez zajmowania kolejnej znaczącej ilości pamięci. Sposób ten może być pomocny w wytwarzaniu kluczy. „ProDNAXORed" jest wynikiem XORingu ProDNA, bez żądania kolejnej znaczącej ilości pamięci. Sposób ten może mieć również znaczenie dla kluczowego wytwarzania.

[0058] Jak pokazano na fig. 3, pary zasad mogą być mapowane na czterobitowy binarny, pozostawiając nadmiarowość do zarządzania niepewnością, gdy są w formie rozpakowanej. Podczas wykonywania obliczeń komputery zwykle używają binarnego (podstawa 2) układu miar. DNA można uznać za podstawę 4,310, ponieważ istnieją 4 formy składowe, a mianowicie G, C, A i T, 305. Mapowanie z podstawy 4 na binarne można przeprowadzić bez nadmiernego obciążenia obliczeniowego, ponieważ każda cyfra o podstawie 4 zajmuje dwie cyfry binarne, chyba że istnieje niepewność co do tego, która cyfra o podstawie 4 jest używana. Taką niepewność można częściowo rozwiązać, traktując DNA jako szesnastkowy (o podstawie 16), a nie o podstawie 4, mapując w ten sposób do 4 bitów na cyfrę szesnastkową.

[0059] Do wytwarzania adresu DNA, w tym szyfrowania i deszyfrowania, chromosomy wymagają przekształcenia w powtarzalną, spójną formę binarną. Te reprezentacje dają spójną formę binarną.

[0060] Szyfrowanie V, które wykorzystuje podstawę czwartą, może oferować największą pewność przy dopasowaniu, ale jest podatne na niepewność spowodowaną niską jakością sekwencjonowania. Dlatego optymalne cele kryptograficzne mogą wymagać użycia skompresowanego ProDNA. Należy zauważyć, że użycie nieskompresowanego ProDNA do szyfrowania V może być nieskuteczne, ponieważ przypadki, w których wkracza korzyść braku kompresji, to również przypadki, w których szyfrowanie V nie ulega odszyfrowaniu. Szyfrowanie U, które wykorzystuje podstawę szesnastą (nazywaną również szesnastkową), oferuje szersze spektrum lub odległość dopasowań i może być uważane za bardziej odporne na niepewność.

[0061] Niepewność może zostać wchłonięta przez większy problem kluczy U, które są tylko częściowymi dopasowaniami do siebie. Klucze V z definicji muszą być do siebie idealnie dopasowane.

[0062] Chociaż niepewność można rozwiązać statystycznie dla celów routingu, jest to problematyczne przy próbie użycia takich „nieprzetworzonych" danych jako współdzielonego tajnego klucza 112 (w co najmniej jednym przykładzie wykonania obejmującym klucz V i/lub mitochondria). Jednak unikalny klucz lub kryptografia głosowania w naturalny sposób rozwiązuje te problemy.

[0063] Podobieństwa są odnotowane między rdzeniem zrzutu komputera a chromosomem. Oba mogą być przeglądane jako seria liczb. Oba programowo definiują algorytmy. W przypadku chromosomów ludzkich programy do budowy człowieka. Rdzeń może wystąpić, gdy komputer ulega awarii. Może

zawierać bałagan liczb z ukrytymi znaczeniami, a także sekwencjami liczbowymi, które nie są już używane lub nigdy nie były używane, ale stanowią dowody. Chromosom może być również rozumiany jako bałagan liczb z ukrytymi znaczeniami, a także sekwencjami liczbowymi, które nie są już używane, ani niektóre nigdy nie były używane, ale stanowią dowody.

[0064] Zarówno zrzut rdzeniowy komputera, jak i chromosom mogą zostać nieumyślnie uszkodzone, zwykle bez lub z minimalnymi negatywnymi konsekwencjami, choć w kontekście chromosomu, bardzo rzadko z pozytywnymi konsekwencjami. Uważa się, że jest to kluczowa siła napędowa ewolucji biologicznej. W kontekście komputerów pozytywne konsekwencje są zwykle projektowane, podczas gdy ewolucja zależy od znacznego wskaźnika ścierania.

[0065] Sposób ProDNA działa w celu zmniejszenia kosztów obliczeniowych podczas analizy zsekwencjonowanego chromosomu, jak opisano w niniejszym dokumencie, przedstawiono na fig. 4. Nieskompresowane ProDNA może na przykład przedstawiać G w postaci binarnej jako 0001. Będzie reprezentować C jako 0010, A jako 0100, a T jako 1000. Skompresowane ProDNA może reprezentować G jako 00, C jako 01, A jako 10 i T jako 11. W tym przykładzie różnica polega na tym, że nieskompresowane zajmuje dwa razy więcej cyfr binarnych niż skompresowane odpowiednie wersje. Odpowiednio, nieskompresowane ProDNA przyjęłoby 1001 jako oznaczające, że istnieje pytanie, czy ta konkretna para zasad to faktycznie G czy T.

[0066] Fig. 4 przedstawia dwa ProDNA 406, z których każdy jest wspierany przez duży magazyn pamięci. Następnie przeprowadzana jest operacja XOR 404, w wyniku czego powstaje trzecie ProDNA 402. To trzecie ProDNA nie jest powieleniem wymagań pamięciowych dwóch oryginalnych, ponieważ łączą je tylko metadane. Rzeczywiste obliczenia są opóźniane do etapu „just in time” (JIT). Może to zapewnić dobrą równowagę między zapotrzebowaniem na obliczenia i pamięć.

[0067] Aby poprawić wiarygodność układu identyfikacji chromosomów, ważne może być odróżnienie wszelkich wyników fałszywie dodatnich od prawdziwych dodatnich. W jednym przykładzie wykonania odbywa się to przez wysłanie zaszyfrowanej wiadomości przy użyciu genomu nadawcy jako klucza. Odbiorca musi mieć wspólne sekwencje DNA z nadawcą, aby odszyfrować wiadomość. Wiadomość będzie niezrozumiała dla innych. Wysyłając wiadomość, układ identyfikacji chromosomów jest w stanie wykryć wymagane informacje, aby przekazać wiadomości z powrotem do nadawcy od innych członków rodziny, niezależnie od tego, skąd zostały wysłane. Wiadomości od innych członków rodziny mogą być następnie dostarczane z powrotem do nadawcy, a tym samym członkowie rodziny ponownie komunikują się ze sobą. Własna poufna sekwencja genomu jednostki jest chroniona przez fakt, że sam adres DNA jest formą skrótu sekwencji genomu jednostki, a nie samą sekwencją. Ta funkcja bezpieczeństwa może powodować ryzyko fałszywych alarmów. Gdy w ten sposób wystąpi fałszywy alarm, można to nazwać dopasowaniem superstanu. Dopasowania superstanowe są rozwiązywane przez nie zakładanie rzeczywistej relacji, dopóki wiadomość nie zostanie pomyślnie zaszyfrowana przy

użyciu dosłownej 140 z krypt 140’ lub unikalnej 150 z krypt w oparciu o sekwencję DNA jednej strony i odszyfrowana przy użyciu sekwencji DNA drugiej strony.

[0068] W co najmniej jednym przykładzie wykonania krypta 605 jest tworzona w formie wielosegmentowej 620. Cała wiadomość jest szyfrowana przy użyciu części klucza, a wynik jest przechowywany w pierwszym segmencie. Następnie cała wiadomość jest ponownie szyfrowana przy użyciu innej części klucza, a wynik jest przechowywany w innym segmencie. Sposób ten można powtarzać wiele razy.

[0069] Szyfrowanie odbywa się przez wykluczanie lub wykluczanie (XOR-ing) każdego odpowiadającego mu bitu w wiadomości z używaną częścią klucza. Ze względu na wydajność, rzeczywista operacja wyłączności lub może być wykonywana w partiach 32-bitowych jako pojedyncza operacja. Wiadomość jest zatem szyfrowana odpowiednią liczbę razy. Odszyfrowanie może nastąpić przy użyciu tego samego klucza, którego użyto do szyfrowania, przez ponowne XOR-owanie odpowiednich części klucza. Jednak nie odszyfrowałoby to przy użyciu sekwencji DNA krewnego, która zazwyczaj zawierałaby wspólne części, ale nie byłaby identyczna.

[0070] Takie odszyfrowanie można osiągnąć w następujący sposób. W jednym przykładzie wykonania deszyfrator tworzy swój klucz w taki sam sposób jak szyfrator, z wyjątkiem tego, że deszyfrator używa własnego zsekwencjonowanego DNA do tworzenia ProDNA, a zatem i klucza. Deszyfrator nie ma dostępu do DNA programu szyfrującego ani nie wie, jaka jest sekwencja programu szyfrującego. Deszyfrator każdy szyfruje każdy segment, XOR sprawdza część klucza. Jeśli istnieją fragmenty DNA wspólne dla obu stron, te fragmenty zostaną poprawnie odszyfrowane. Tam, gdzie ich nie ma, te wartości bitów będą miały średnio 50:50 szans na poprawność lub błędność.

[0071] Tam, gdzie ten sam bit jest zaszyfrowany w wielu segmentach, a program szyfrujący i deszyfrujący współdzielą połowę swojego DNA, a szyfrowany bit to 1, oczekuje się, że wynikną następujące wartości:

50% = typowy procent czasu, w którym jest przechowywany jako 0 w krypcie.

50% = typowy procent czasu, w którym jest przechowywany jako 1 w krypcie.

0% = procent czasu, w którym XOR odszyfrowuje z powrotem do 0, gdzie zarówno nadawca, jak i odbiorca dzielą odpowiedni bit w zsekwencjonowanym DNA.

100% = procent czasu, w którym XOR odszyfrowuje z powrotem do 1, gdzie zarówno nadawca, jak i odbiorca dzielą odpowiedni bit w zsekwencjonowanym DNA.

50% = procent czasu, w którym XOR odszyfrowuje z powrotem do 0, gdzie zarówno nadawca, jak i odbiorca nie dzielą odpowiedniego bitu w zsekwencjonowanym DNA.

< 50% = całkowity procent czasu, w którym XOR odszyfrowuje z powrotem do 0.

> 50% = całkowity procent czasu, w którym XOR odszyfrowuje z powrotem do 1.

[0072] Odszyfrowanie można osiągnąć przez głosowanie etap po etapie, a następnie testowanie wyniku względem wartości kontrolnej, jak opisano poniżej. Może istnieć kilka czynników kalibracyjnych, które mogą służyć zwiększeniu lub zmniejszeniu prawdopodobieństwa udanego odszyfrowania, a tym samym wykryciu, jak blisko spokrewnione muszą być osoby, aby uzyskać powiedzenie. Można je ustawić jako proporcję ustawienia głośności.

[0073] Utrzymanie dużych scentralizowanych kolejek sekwencyjnych może być trudne. Jednak w co najmniej jednym przykładzie wykonania jakiekolwiek zsekwencjonowane kolejki generowane przez układ identyfikacji chromosomów nie muszą być scentralizowane. Wiadomości wymagają tylko przekierowania do właściwego miejsca docelowego. Przełączanie pakietów TCP/IP już pokazało, że takie problemy można rozwiązać globalnie. Kolejki zsekwencjonowane są używane do trasowania, a zatem jeden lub więcej algorytmów może być użytych do trasowania przy użyciu kolejek.

[0074] Układ identyfikacji chromosomów umożliwia przechowywanie rozproszonej bazy danych wiadomości 656 w rozproszonej sieci węzłów, w tym adresie DNA. Adres DNA to adres nadawcy, ale także nadawca i odbiorca albo dzielą fragment swoich odpowiednich adresów DNA, albo odpowiednia część ich adresów DNA zawiera jedną lub więcej liczb, które mają z góry określoną bliskość względem siebie. Każdy węzeł wysyła każdą wiadomość, którą otrzymuje klasycznie, do jednego lub większej liczby innych węzłów. Węzeł wysyła wiadomość do węzłów, które zgodnie ze znanymi danymi mają wiadomości przychodzące lub przechodzące do tych fragmentów adresu. Ta technika przetwarzania rozproszonego prowadzi do tego, że wiadomości bliskich krewnych spotykają się ze sobą, zbiegają się, a tym samym docierają do tych samych węzłów.

[0075] Sposób ten jest wykonywany przez przejrzenie 12 wartości Σ z tabeli 105 w adresie, z którego wiadomość pochodzi i/lub do którego zmierza. W każdym z tych 12 przypadków znajduje się poprzednia wiadomość, która miała ten sam numer w tym samym miejscu lub, jeśli nie została znaleziona, dwa najbliższe dopasowania opcjonalnie zawierające numer powyżej i numer poniżej. W jednym przykładzie wykonania może to zapewnić do 24 innych wiadomości. Lista ta jest następnie rozszerzana, aby uwzględnić inne wiadomości, które zostały powiązane jako część potencjalnych rodzin tych do 24 wiadomości. Kilka z tych wiadomości i potencjalnych rodzin mogło zostać policzonych podwójnie, a co za tym idzie, zredukowanych do jednej liczby. Wiadomość, którą węzeł próbuje przesłać dalej, zostanie następnie przesłana na wszystkie adresy internetowe, które były skojarzone z tymi wiadomościami po tym, jak same adresy internetowe zostaną zredukowane do pojedynczej liczby.

[0076] Adres DNA może zawierać jedną znaczącą liczbę dodaną. Biorąc pod uwagę 12 liczb już wspomnianych, możliwe jest, że dwie osoby z tej samej rodziny będą miały wszystkie 12 liczb ujawnionych w niniejszym dokumencie, zawierających identyczne wartości. Ta trzynasta liczba służy

do odróżnienia członków rodziny od siebie i może być reprezentowana jako skrót całego ich genomu. Jednak nadal niesie to ryzyko pomylenia identycznego rodzeństwa. Tam, gdzie uważa się, że należy uniknąć takiego ryzyka, że istnieje identyczne rodzeństwo, na hash można nałożyć dodatkowy znacznik czasu tworzenia sekwencji.

[0077] Własny węzeł danej osoby jest jedynym węzłem, który ma dostęp do sekwencji genomu danej osoby. Służy do tworzenia ich adresu DNA, szyfrowania wiadomości, które wysyłają i próby odszyfrowania otrzymywanych wiadomości.

[0078] Problem z adresami superpozycji może powstać, gdy dwie niespokrewnione osoby mogą wydawać się spokrewnione, jako efekt uboczny wykorzystania informacji o routingu adresów, których nie można wykorzystać do inżynierii wstecznej genomu danej osoby. Mechanizm szyfrowania zapewnia, że tylko powiązane osoby będą mogły czytać wiadomości drugiej strony. Dlatego układ identyfikacji chromosomów ma na celu rozwiązanie problemu oszustów i superpozycji DNA przez sprawdzenie, czy odszyfrowanie zakończyło się sukcesem, aby ustalić, czy osoby są rzeczywiście spokrewnione. Tam, gdzie pokrewieństwo jest w ten sposób ustalone, jednostki są następnie łączone w jednostki rodzinne, ponieważ gdy osoby A i B są spokrewnione z osobą C, to osoby A i B są również spokrewnione ze sobą. Niepowodzenie odszyfrowania może być ocenione jako wskazujące na oszusta lub przypadek superpozycji adresów DNA.

[0079] Układ identyfikacji chromosomów może być zorganizowany tak, aby zawierał mechanizm „objętości". W miarę zwiększania miary objętości, dalsze związki genetyczne będą akceptowane jako powiązane.

[0080] Aby wysłać wiadomość do członków rodziny, może być konieczne wysłanie (lub „przekierowanie") wiadomości od osoby na adres jednego lub więcej członków jej rodziny, przy czym znana jest tylko sekwencja genomu nadawcy**.** W centrum adresu znajduje się kilka liczb opisowych. Adresy są istotne względem siebie. Można obliczyć odległość między dowolnymi dwoma adresami. Liczby opisowe są indeksowane w taki sposób, że można również łatwo uzyskać zbliżone wartości.

[0081] Każdy węzeł chromosomalnego układu identyfikacji może wysyłać wiadomości do najbliższych znanych krewnych, powodując w ten sposób występowanie klastrów.

[0082] Można stosować dwie różne formy szyfrowania:

1) Verbatim 140 - przy użyciu dowolnego z mitochondriów 115 lub chromosomów X 120 lub Y 125. W oparciu o wspólne i nieujawnione publicznie klucze kryptograficzne, dosłowny klucz kryptograficzny 140 wykorzystuje fakt, że te trzy chromosomy mogą być przekazywane z rodzica na potomstwo, z pewnymi zastrzeżeniami.

2) Unique 150 - z wykorzystaniem chromosomów o numerach od 1 do 22, jak pokazano w punkcie 155.

[0083] Te chromosomy razem, z rzadkimi wyjątkami, takimi jak w przypadku identycznego rodzeństwa, są konwencjonalnie akceptowane jako unikalne dla każdego osobnika. Szyfrowanie odbywa się za pomocą chromosomów, które zazwyczaj są unikalne dla każdego osobnika, ale biorąc pod uwagę fakt, że fragmenty wyżej wymienionych chromosomów mogą być dzielone z krewnymi lub nawet tylko z innymi istotami ludzkimi. Wspólna proporcja może wskazywać na biologiczną bliskość dwóch osób. Odszyfrowanie wymaga schematu głosowania, biorąc pod uwagę fakt, że ponad 50% klucza dostępnego do odszyfrowania może zostać „złamane" z punktu widzenia kryptografa. W konwencjonalnej kryptografii odbiorca kodu polega na posiadaniu pełnego i poprawnego klucza w celu odszyfrowania. W tym kontekście, ponieważ klucz do szyfrowania i inny klucz do odszyfrowania są pobierane od różnych osób, które dzielą tylko część swojego DNA, niewspółdzielone części można uznać za zepsute lub po prostu biały szum z punktu widzenia odszyfrowywania.

[0084] Następnie można przeprowadzić kontrolę weryfikacyjną, wykorzystując fakt, że odszyfrowanie jest nawet możliwe. O ile klucz, zawierający dane z materiału genetycznego nadawcy, nie jest udostępniany we wcześniej ustalonym stopniu podobieństwa, odszyfrowanie nie może nastąpić. Dlatego udane odszyfrowanie wskazuje na z góry określony stopień biologicznego spokrewnienia. Mechanizm cyfr kontrolnych lub „wartość kontrolna" może być dołączony do wiadomości i w ten sposób użyty do wykonania takiego określenia. Zmniejsza to koincydencję powtarzających się części wiadomości zbieżnych z potencjałem wszelkich powtarzających się części klucza. W przykładzie klucza z dwoma (skompresowanymi) lub czterema (nieskompresowanymi) bitami na parę zasad, wspólne pary zasad często pokrywają się ze wspólnymi znakami. Dlatego też, gdyby zestaw znaków używał liczby bitów innej niż liczba pierwsza, na przykład osiem, problem można przezwyciężyć przez zastosowanie obszaru typu blob słów o rozmiarze liczby pierwszych (BoPSW) 652, jak szczegółowo opisano poniżej.

[0085] Przy określaniu z góry określonego stopnia biologicznej bliskości rodziny najbliżsi krewni według jednej definicji będą dzielić identyczną kopię mitochondriów 115, chromosomu X 120 i/lub chromosomu Y 125. W tych przypadkach szyfrowanie i deszyfrowanie uzyskuje się za pomocą wspólnej tajnej kryptografii 112, 150, 102' lub „kryptografii klucza dosłownego" 140, 140', jak opisano w niniejszym dokumencie, gdzie klucze pochodzą z mitochondriów 115, chromosomów X lub Y 120, 125, jak opisano powyżej. W przypadkach, w których ludzie mogą być spokrewnieni, ale nie tak blisko, lub występują błędy w sekwencjonowaniu lub niepewność sekwencjonowania, klucze pochodzą z chromosomów od 1 do 22, 155, przy czym zaszyfrowana wiadomość 202 zawiera wiele kopii wiadomości, zaszyfrowanych różnymi częściami klucza. Zastosowany zespół kryptowy może być określany jako wiązka kryptowa 605. Wiele wiadomości 208 może być zawartych z wybranych członków rodziny w pojedynczej wiadomości 202. Wielokrotnie podejmowana jest próba

odszyfrowania, a mechanizm głosowania 165 jest używany do określenia, jaką wartość powinien mieć każdy bit. Tam, gdzie nie ma pełnej kopii chromosomu umożliwiającej współdzielenie tajnej kryptografii, istnieje potrzeba szyfrowania i odszyfrowywania, w której nadawca i powiązany odbiorca mają klucze, których części współdzielą, których części nie mają i nie wiedzieć, które porcje są które. Jest to „kryptografia z unikalnym kluczem" lub „kryptografia głosowania", jak dalej opisano w niniejszym dokumencie. Klucz może zostać wygenerowany przez:

- digitalizacja sekwencji DNA nadawcy;
- klasycznie formatując go w formie nieskompresowanej, ale można również użyć innych form;
- nakładanie przez operacje XOR różnych chromosomów;
- składanie chromosomów z powrotem na siebie opcjonalnie przy użyciu XOR; i/lub
- przez różne inne operacje matematyczne.

[0086] Generowanie indeksu posortowanego według adresu genetycznego DNA i akceptowanie najbliższego dostępnego, a tym samym użycie jednego lub więcej wyników przetwarzania rozproszonego, może znacznie skrócić czas potrzebny na znalezienie krewnego w porównaniu z konwencjonalnymi sposobami wyszukiwania. Jednak możliwe jest, że niektóre osoby nie chcą być odnalezione, a zatem mogą zostać odnalezione tylko wtedy, gdy zdecydują się na użycie układu identyfikacji chromosomów, nawet jeśli faktycznymi dostawcami układu są różne jednostki. Dlatego najbliżsi krewni korzystający z układu identyfikacji chromosomów mogą być połączeni. Osoby niebędące użytkownikami nie będą połączone, a więc mogą pozostać w ukryciu przed swoimi bliskimi za pomocą aranżacji.

[0087] Może być dostarczone jedno lub więcej narzędzi kalibracyjnych, służących do regulacji, jak daleko mogą wystąpić dopasowania. Można to zredukować, aby zapewnić większą łatwość użycia w jednej metryce zwanej „głośnością". „Najgłośniejszy" poziom głośności, który można ustawić na maksymalną wartość 100, można opisać jako transmisję, w której można wykryć bardziej odległe relacje biologiczne. Głośność 1 może być raczej „szeptem", tak że można znaleźć tylko bardzo bliskich krewnych. Obliczenia można zastosować do adresów DNA dwóch różnych osób, co daje liczbę całkowitą reprezentującą najkrótszą możliwą odległość genetyczną między tymi dwiema osobami w kategoriach par zasad. Użytkownik może dodatkowo chcieć „wyciszyć" komunikaty od innych osób o wysokim poziomie głośności, aby uniknąć przeszkadzania wiadomościami, których odbieraniem nie jest zainteresowany.

[0088] Jak pokazano na fig. 5, można zauważyć, że niektóre prawne lub pośrednie relacje, takie jak mąż 505 i żona 510, mogą nie dzielić znacznej części DNA, ale nadal mogą chcieć ponownie się połączyć. W tym przykładzie mąż 505 może być określany jako ojciec i odpowiednio żona 510 może być

określana jako matka. Tacy użytkownicy mogą nadal znajdować się nawzajem przy użyciu układu identyfikacji chromosomów, pod warunkiem, że istnieje pośrednik, taki jak ich wspólne biologiczne potomstwo 515, korzystające z układu. Potomstwo, biologicznie spokrewnione z każdym użytkownikiem, może pełnić funkcję pomostu łączącego tych dwóch użytkowników pomimo braku biologicznego podobieństwa. W przeciwieństwie do każdego z ich związków z potomstwem, mąż 505 i żona 510 nie mają ze sobą wspólnego przodka. Aby dwóch użytkowników mogło być bezpośrednio połączonych jako spokrewnieni, wymagane jest wzajemne połączenie rodzinnego DNA.

[0089] Każdy węzeł, podczas odszyfrowywania wiadomości, może działać jako przekaźnik, tym samym ponownie szyfrując wiadomość przy użyciu własnego klucza pośredniczącego. Może to być operacja „wieloskokowa", zwiększając w ten sposób zakres wyszukiwania. Wynikająca z tego zbieżność wiadomości, ujawniony tutaj mechanizm trasowania, umożliwia węzłom łączenie potencjalnych rodzin. Gdy A jest powiązane z B, a B jest powiązane z C, oznacza to, że A jest powiązane z C. Jeżeli węzeł B pomyślnie przełoży wiadomość od A, może być zdolny do ponownego wysłania wiadomości i tym razem umożliwienia C odszyfrowania, a tym samym przejrzenia wiadomości. Ta retransmisja jest wydawana pod nowym numerem wiadomości, ale zawiera wersję oryginalną. Może to skutkować wieloma różnymi zaszyfrowanymi kryptami tej samej wiadomości, umożliwiając różne próby odszyfrowania w dowolnym węźle odbiorczym.

[0090] Podczas gdy adres genetyczny DNA danej osoby pochodzi z sekwencji genomu danej osoby, z założenia nie jest możliwe odtworzenie sekwencji genomu na podstawie jej adresu. W ten sposób sekwencja genomu jest utrzymywana w tajemnicy. Chociaż szyfrowanie i deszyfrowanie może wykorzystywać sekwencję genomu danej osoby, sama sekwencja nie jest przesyłana i dlatego pozostaje tajna.

[0091] Testowanie każdej osoby przeszukującej względem każdej innej osoby przeszukującej może generować niewykonalnie dużą liczbę wyszukiwań do osiągnięcia. Jeśli populacja Ziemi wynosi 7 600 000 000, wymagałoby to porównania 7 600 000 000 testów DNA z 7 600 000 000 innych testów DNA, co daje łącznie 57 760 000 000 000 000 000 porównań. Rzeczywiście, biorąc pod uwagę czas wymagany do wykonania konwencjonalnego testu DNA, ludzie mogą być dodawani do listy szybciej, niż mogą zostać przetestowani. W tym kontekście można to przezwyciężyć, stosując ogólny adres DNA, aby znacznie ograniczyć opcje, a następnie stosując kryptografię, ponieważ wiadomości można odszyfrować tylko wtedy, gdy są naprawdę powiązane. Co więcej, układ identyfikacji chromosomów może być zorganizowany tak, aby wykorzystać przetwarzanie rozproszone, ponieważ każdy użytkownik może mieć swój własny węzeł z własną sekwencją DNA. Węzły te mogą obejmować na przykład mobilny smartfon. Inne środki implementacji mogą obejmować jeden lub więcej spośród: mobilnej aplikacji na smartfona („aplikacja"); serwer obsługujący wspomnianą aplikację mobilną na smartfona; wersja aplikacji mobilnej na smartfona na komputer PC lub laptop; bibliotekę

oprogramowania do tworzenia aplikacji mobilnych na smartfony, serwery obsługujące aplikacje mobilne na smartfony i/lub wersję mobilnej aplikacji na smartfony/na komputer PC i/lub laptop; wszelkie inne urządzenia sieciowe; usługę poczty tekstowej, e-mail lub poczty głosowej, w której numer lub adres pochodzą z własnego profilu DNA użytkownika; i/lub usługę transmisji, gdzie informacje medyczne mogą być interesujące dla określonych rodzin. Przewodowe połączenie sieciowe, takie jakie można znaleźć w serwerze domowym 204, w przeciwieństwie do mobilnego smartfona, może zapewniać korzyść w postaci bardziej spójnego sygnału lub połączenia z siecią zewnętrzną, taką jak Internet, a także wnęką do przechowywania/zapisywania wiadomości przychodzących, gdy smartfon może być poza zasięgiem lub odłączony od sieci zewnętrznej. Ponadto skrót serwera domowego zawierający adres IP może służyć do rozróżniania identycznych bliźniaków, ponieważ każdy z nich prawdopodobnie miałby inny adres IP.

[0092] W jednym przykładzie wykonania szyfrowanie odbywa się za pomocą bloku słów o rozmiarze pierwszym (BoPSW). „Blob" konwencjonalnie dotyczy dużego obiektu binarnego. Wiadomości do zaszyfrowania są umieszczane w BoPSW. BoPSW przechowuje znaki lub słowa o określonej wielkości w bitach, które są liczbą pierwszą. Umożliwia to odwzorowanie szeregu kodów znaków do standardowej postaci bez wprowadzania słabości kryptograficznej zbieżności danych i słów kluczowych. Zrozumiałe jest, że ten przykład ilustruje częściowe odszyfrowanie, a takie odszyfrowanie może być wykonywane na zasadzie bit po bicie, a nie znak po znaku. Komputery mają zwykle rozmiary słów 8, 16, 32 lub 64 bity. ASCII, początkowo zdefiniowane w siedmiu bitach, ale zwykle używane w postaci 8-bitowej, jest zbyt ograniczone do celów międzynarodowych. Unicode, z jego 16 bitami, może być uważany za zapewniający większą internacjonalizację niż ASCII. Istnieją również inne zestawy znaków, na przykład specjalnie zaprojektowane dla znaków chińskich lub japońskich. Celem ujawnionego tutaj układu jest niewykluczanie opcji ustawień regionalnych, przy jednoczesnym zachowaniu poprawności kryptograficznej. Mapowania znaków jeden do jednego nie są konwencjonalnie uważane za poprawne kryptograficznie, nawet jeśli zostały przypadkowo wprowadzone. Ryzyko przypadkowego wprowadzenia jest zmniejszone, gdy rozmiary słów danych i klucza nie pokrywają się. BoPSW przechowuje wiadomości w postaciach, które prawdopodobnie nie pokrywają się z naturalnymi ustawieniami DNA. Potencjalne rozmiary słów używane przez BoPSW to na przykład liczby pierwsze; 7, 11, 13, 17, 19, 23, 29 i 31. Zrozumiałe jest, że większe liczby pierwsze mogą być użyte zamiast lub w uzupełnieniu podanych przykładów.

[0093] Fig. 6a, 6b i 6c przedstawiają jeden konkretny przykład wykonania dla nominalnego użytkownika „Ewa" 600, jej męża „Adam" 610 i jej syna „Abla" 615. Adam nie jest krewnym Ewy. DNA Ewy jest zsekwencjonowane 650, a adres DNA można przedstawić w sposób przedstawiony powyżej w tabeli 105:

	#			
	G	C	A	T
M	#	#	#	#
X	#	#	#	#
X/Y	#	#	#	#

[0094] W powyższej tabeli poczyniono następujące odniesienia:

„M” dotyczy chromosomu mitochondrialnego osobnika;

„X” dotyczy jednego z chromosomów X osobnika;

„X/Y” dotyczy drugiego chromosomu X w przypadku kobiety lub chromosomu Y w przypadku mężczyzny;

„G” dotyczy par zasad GC;

„C” dotyczy par zasad CG;

„A” dotyczy par zasad AT; oraz

„T” dotyczy par zasad TA.

[0095] Zauważono powyżej, że biologiczne samice na ogół mają dwa chromosomy X, podczas gdy biologiczne samce na ogół mają tylko jeden.

[0096] Każdy symbol „#” reprezentuje liczbę. Poza # w najwyższym rzędzie, każdy # jest obliczany przez zliczenie liczby wystąpień pary zasad, której dotyczy, w odpowiednim sekwencjonowanym chromosomie. Jak powyżej, obliczona liczba jest następnie mnożona przez współczynnik ufności, zwany również współczynnikiem pewności, który w tym przykładzie wykonania wynosi 12.

[0097] Czasami istnieje pewien stopień niepewności, która para zasad jest prawidłowa. Gdy taka niepewność występuje, na przykład ponieważ istnieje wielu potencjalnych kandydatów 654, wykonuje się obliczenia, w których liczbę potencjalnych kandydatów mnoży się przez liczbę w następujący sposób:

Potencjalni Kandydaci	Pomnożyć przez
1	12
2	6
3	4
4	3

[0098] W przypadku wielu kandydatów, każdy kandydat powinien mieć zastosowany współczynnik zaufania. W rezultacie każda para zasad w sekwencji spowoduje całkowity wzrost dwunastokrotnie.

[0099] Te # wartości są następnie wykorzystywane do kierowania wiadomości do miejsca docelowego.

[0100] # w najwyższym wierszu powyższej tabeli reprezentuje kod skrótu całej sekwencji genomu osobnika. Opcjonalnie, gdy istnieje ryzyko identycznego rodzeństwa, jako część tego kodu skrótu można również dołączyć znacznik czasu epoki reprezentujący datę i godzinę sekwencjonowania.

[0101] Następujące cechy algorytmu mieszającego tego najwyższego # mogą obejmować:

- hash całego zsekwencjonowanego chromosomu;
- spójność we wszystkich wdrożeniach; oraz
- zgodność z klasycznymi zasadami wartości hash.

[0102] Ta najwyższa wartość # jest używana w tym przykładzie wykonania, aby odróżnić nadawcę od innych nadawców. Każda z pozostałych wartości może być podzielana przez inne osoby. W pewnych okolicznościach wszystkie dwanaście z # wartości może być identyczne.

[0103] Specjalne działanie może zostać podjęte, jeśli węzeł jest tak świeży, że nie ma danych ani wiedzy o innych wiadomościach, które przeszły przez węzeł. W takich okolicznościach przekazuje wiadomość w górę do wcześniej zdefiniowanej bramy, która będzie w stanie przeprowadzić przetwarzanie, jak opisano poniżej.

[0104] Każdy z dwunastu wątków przyjmuje jedną z dwunastu wartości #. Każdy wątek porównuje swoją wartość # z ostatnim wykryciem wiadomości z adresem o tym samym numerze w tym samym miejscu w tabeli adresów. Jeśli ich nie było, co może mieć miejsce, używa najbliższej liczby powyżej i najbliższej poniżej, tak jakby były dopasowanie. Zaszyfrowana wiadomość zaadresowana DNA zostanie następnie przesłana na adres, o którym wcześniej wiadomo było, że ma wiedzę o tym numerze. Gdy wiele wątków próbowałoby wysłać do tego samego innego węzła, transmisja odbywa się tylko raz.

[0105] Każdy węzeł, który odbierze wiadomość, przeprowadzi ten sam sposób przekazywania, chociaż zapobiega się pętlom zwrotnym. Mechanizm ten napędza bruzdy specyficzne dla elementu adresowego DNA w globalnej sieci powiązanych węzłów zgodnych. Tam, gdzie przecinają się bruzdy, relacje biologiczne mogą zostać ponownie połączone. Każdy adres DNA ma dwanaście poziomów, do których zmierza. Dwunastka składa się z trzech zestawów po cztery, a klasycznie wszystkie cztery w danym zestawie będą pasować w tych samych czterech zestawach węzłów. Pozostałe dwa zestawy prawdopodobnie nie będą pasować w tych samych węzłach, ponieważ będą sięgać do innej gałęzi rodziny.

[0106] Każdy węzeł podejmie próbę odszyfrowania, aby odkryć, czy wiadomość jest przeznaczona dla właściciela węzła. Fig. 6a, 6b i 6c ilustrują przejście wiadomości do odbiorcy.

[0107] W odniesieniu najpierw do fig. 6a, przykładowy użytkownik „Ewa” 600 ustawia głośność na niską wartość, dążąc do zidentyfikowania tylko bardzo bliskich krewnych – tj. tych o bardzo podobnym

DNA - zaczynając od głośności 1 z maksymalnej wartości 100 w tym przykładzie wykonania. Za każdym razem, gdy Ewa korzysta z układu, może na przykład zwiększyć głośność, aby zwiększyć zasięg/odbiorcy, do których dociera jej wiadomość.

[0108] Ewa zsekwencjonowała swoje DNA i ma dane o sekwencji DNA 650. Ewa napisała również wiadomość 100 do przekazania wszystkim bliskim krewnym: „Droga rodzino, jestem tutaj, gdzie jesteś? Kocham, Ewa".

[0109] Jak omówiono w opisanych tutaj przykładach wykonania, wiadomość jest ponownie sformatowana do postaci blobu słów o rozmiarze pierwszym (BoPSW) 652.

[0110] Korzystając z danych sekwencji DNA 650 Ewa, tworzony jest klucz U 150, a także kilka kluczy V 140 i adres DNA dla Ewy 105.

[0111] Klucz U 150 i BoPSW 652 są używane do szyfrowania wiadomości przy użyciu klucza U 150, podczas gdy klucze V 140 są używane do tworzenia kilku zaszyfrowanych wersji BoPSW 652. Wartości kontrolne są tworzone obok zaszyfrowanej wiadomości wielosegmentowej U kryptu 620, a także względem V kryptu 140', która zawierała wiele zaszyfrowanych wersji wiadomości 100, które są zaszyfrowane każdym z kluczy V 140.

[0112] Adres komputera domowego Ewy (lub serwera domowego, a potencjalnie nawet zdalnego serwera) 204 jest dodawany wraz z wiadomością wielosegmentową U kryptu 620 (i sprawdza wartość) i V kryptu 140' (i sprawdza wartość) oraz adresem DNA 105 w wiązkę kryptu 605.

[0113] Wiązka kryptu 605 może być następnie przesyłana w 12 oddzielnych częściach, po jednej na element układu adresu DNA.

[0114] W odniesieniu teraz do fig. 6b, inny użytkownik otrzymuje wiązkę kryptu Ewy 605 i albo przesyła dalej, albo próbuje odczytać wiadomość wewnątrz lub jedno i drugie.

[0115] Jeśli wiązka kryptu jest przekazywana, można ją zaadresować na podstawie poprzedniego ruchu, który przeszedł przez węzeł, aby określić, do którego innego węzła wysłać wiązkę kryptu i/lub przesłać do jednego lub więcej innych węzłów przy użyciu adresu DNA Ewy 105.

[0116] Aby odczytać wiadomość Ewy, Abel używa ich sekwencji DNA 615, z której utworzono klucz U-key 150 i V-Keys 140/v-vault 140' oraz adres DNA 105. Za pomocą tych kluczy tworzonych jest kilku kandydatów 654 dla odszyfrowanej wiadomości (wraz z wartością kontrolną) i są one określane jako czytelne lub nie. Jeśli jest czytelny, to wiadomość 100 może być pokazana Abelowi, ale jeśli nie jest czytelny, oznacza to, że Ewa nie jest wystarczająco bliską krewną w oparciu o głośność jej wiadomości, aby Abel mógł przeczytać jej wiadomość 100. Jeśli jednak ustawienia Abla dotyczące odbioru wiadomości są ustawione na próg niższy niż głośność Ewy, wówczas Abel nie odczyta wiadomości Ewy 100, niezależnie od tego, czy jest czytelna.

[0117] W oparciu o deszyfrowanie w węźle Abla, wiadomość Ewy 100 może zostać przekazana do innego węzła, jak na fig. 6c, do której teraz będziemy się odnosić, ale ponownie zakodowana przy użyciu kluczy pochodzących z DNA Abla 140,160 i adresu DNA 105 w celu ponownej transmisji do innych członków rodziny zgodnie z ustawieniami głośności Ewy – co w tym przypadku ma zastosowanie, ale jeśli ustawienia głośności nie mają zastosowania, ponowna transmisja nie nastąpi.

[0118] Na fig. 6c Adam odbiera wiadomość Ewy zaszyfrowaną kluczami Abla jako pakiet kryptowy 605 i powtarza próby odszyfrowania wiadomości 100 od Ewy przy użyciu własnych kluczy v 140 Adama i klucza u 150 do odszyfrowania szyfrowania Abla. W tym przypadku Adam jest na przykład mężem Ewy, więc nie jest krewnym Ewy, ale krewnym Abla, więc będzie mógł odszyfrować wiadomość Ewy przez Abla zgodnie z ustawieniami głośności Ewy.

[0119] W każdym przypadku wiadomości mogą być przekazywane Ewie, że ktoś odszyfrował jej wiadomość.

[0120] W alternatywnym przykładzie wykonania zapewniono bezpieczny układ przesyłania wiadomości. Układ przesyłania wiadomości może wykorzystywać jedną lub więcej technik szyfrowania i/lub deszyfrowania opisanych w odniesieniu do jednego lub więcej innych przykładów wykonania. Na przykład klucze szyfrowania i deszyfrowania mogą być generowane dla użytkownika i udostępniane w razie potrzeby, aby umożliwić innym użytkownikom bezpieczne przesyłanie wiadomości do tego użytkownika. Podobnie wielu użytkowników może odpowiednio współdzielić takie klucze, aby umożliwić komunikację dwukierunkową lub wielokierunkową.

[0121] W kolejnym alternatywnym przykładzie wykonania zapewniony jest sposób weryfikacji tożsamości. Sposób weryfikacji obejmuje użycie jednej lub więcej technik szyfrowania i/lub deszyfrowania opisanych w odniesieniu do jednego lub więcej innych przykładów wykonania i może być połączony z wykorzystaniem jednej lub więcej innych technik weryfikacji tożsamości i/lub biometrycznych.

[0122] Dowolna cecha układu, jak opisano w niniejszym dokumencie, może być również dostarczona jako cecha sposobu i odwrotnie. W znaczeniu tutaj użytym, środki plus cechy funkcji mogą być wyrażane alternatywnie w kategoriach ich odpowiedniej struktury.

[0123] Każda cecha w jednym aspekcie może być zastosowana w innych aspektach, w dowolnej odpowiedniej kombinacji. W szczególności aspekty sposobu mogą być zastosowane do aspektów układu i odwrotnie. Ponadto dowolne, niektóre i/lub wszystkie cechy w jednym aspekcie można zastosować do dowolnych, niektórych i/lub wszystkich cech w dowolnym innym aspekcie, w dowolnej odpowiedniej kombinacji.

[0124] Należy również zauważyć, że poszczególne kombinacje różnych cech opisanych i zdefiniowanych w dowolnych aspektach mogą być wdrażane i/lub dostarczane i/lub używane niezależnie.

Zastrzeżenia patentowe

1. Wspomagany komputerowo sposób lokalizowania jednego lub więcej członków sieci rodzinnej, obejmujący etapy:

generowanie jednego lub większej liczby kluczy szyfrowania (140, 150) pochodzących z pierwszej sekwencji genomowej (650);

szyfrowanie wiadomości (100) przy użyciu jednego lub każdego klucza szyfrowania (140, 150) w celu utworzenia zaszyfrowanej wiadomości;

wysyłanie zaszyfrowanej wiadomości do jednego lub większej liczby zdalnych urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie zdalnych urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania (140, 150) pochodzących z drugiej sekwencji genomowej; oraz

otrzymanie potwierdzenia, czy odszyfrowanie zaszyfrowanej wiadomości zakończyło się powodzeniem przez jedno lub więcej zdalnych urządzeń;

przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej (650).

2. Sposób według zastrz. 1, obejmujący ponadto etapy:

odbierania danych wejściowych zawierających pierwszą sekwencję genomową (650); oraz

generowania jednego lub więcej kluczy szyfrowania (140, 150) na podstawie danych wejściowych.

3. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym etap odszyfrowania zaszyfrowanej wiadomości na jednym lub większej liczbie zdalnych urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania (140, 150) pochodzących z jednej lub większej liczby dalszych sekwencji genomowych.

4. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym jeden lub więcej kluczy szyfrowania (140, 150) zawiera: pierwszy klucz szyfrowania (140) dla algorytmu szyfrowania dosłownego i drugi klucz szyfrowania (150) dla unikalnego algorytmu szyfrowania;

opcjonalnie przy czym pierwszy klucz szyfrowania (140, 150) jest używany w odniesieniu do: jednego lub większej liczby mitochondriów; pierwszy chromosom X; i drugi chromosom X lub chromosom Y,

utworzenia trzy dosłownie algorytmicznie generowane szyfry w odniesieniu do pierwszej sekwencji genomowej (650),

ponadto opcjonalnie przy czym drugi klucz szyfrowania (140, 150) jest używany do szyfrowania wiadomości przy użyciu kombinacji chromosomów od 1 do 22 zawartych w pierwszej sekwencji genomowej (650).

5. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym etap wysyłania zaszyfrowanej wiadomości jest wykonywany przy użyciu jednego lub większej liczby danych wyjściowych pochodzących z pierwszej sekwencji genomowej (650).

6. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym pomyślne odszyfrowanie zaszyfrowanej wiadomości wskazuje na z góry określony poziom spokrewnienia.

7. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym dowolna lub dowolna kombinacja:

pierwszej i/lub drugiej sekwencji genomowej zawiera co najmniej część sekwencji genomu; i/lub

przy czym sekwencja genomowa (650) zawiera cztery zasady zawierające jedną lub więcej spośród: guaniny (G), cytozyny (C), adeniny (A), tyminy (T) i/lub uracylu (U);

opcjonalnie przy czym cztery zasady są mapowane na sekwencję binarną;

ponadto opcjonalnie przy czym sekwencja binarna zawiera dwie formy: skompresowaną i/lub nieskompresowaną; i/lub

zastrzeżeń wejście zawiera liczbę jednej lub więcej z czterech zasad.

8. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym odszyfrowanie zaszyfrowanej wiadomości kończy się pomyślnie tylko wtedy, gdy z góry określona proporcja pierwszej sekwencji genomowej (650) odpowiada drugiej sekwencji genomowej, opcjonalnie zastrzeżeń z góry określona proporcja odzwierciedla poziom pokrewieństwa między właścicielem pierwszej sekwencji genomowej (650) i właścicielem drugiej sekwencji genomowej.

9. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym etap wysyłania zaszyfrowanej wiadomości obejmuje utworzenie układu liczb całkowitych trzy na cztery i unikalnego indywidualnego zaszyfrowanego identyfikatora liczb całkowitych; opcjonalnie przy czym unikalna pojedyncza zaszyfrowana liczba całkowita jest skrótem sekwencji całego genomu (650); przy czym ponadto opcjonalnie dodatkowo zawiera nałożenie znacznika czasu daty sekwencjonowania.

10. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym etap szyfrowania wiadomości obejmuje użycie binarnego dużego obiektu słów o pierwszej wielkości (BoPSW) (652).

11. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym dowolna lub dowolna kombinacja:

jednego lub więcej węzłów rozproszonej sieci węzłów jest zdolna do wysłania zaszyfrowanej wiadomości tylko za pierwszym razem, gdy zaszyfrowana wiadomość zostanie odebrana przez ten lub każdy węzeł; i/lub

każdy węzeł sieci rozproszonej ma skojarzonego użytkownika.

12. Sposób według któregokolwiek z poprzednich zastrzeżeń, obejmujący ponadto etap: wyprowadzania miary odległości genetycznej między pierwszą sekwencją genomową (650) a drugą sekwencją genomową; opcjonalnie przy czym miara odległości genetycznej jest określana przy użyciu adresu DNA (105).

13. Sposób według któregokolwiek z poprzednich zastrzeżeń, przy czym zaszyfrowana wiadomość jest przechowywana przez pewien czas na zdalnym serwerze, opcjonalnie przy czym odszyfrowanie zaszyfrowanej wiadomości po etapie wysłania zaszyfrowanej wiadomości odbywa się z opóźnieniem czasowym.

14. Urządzenie do lokalizacji sieci rodzinnej, zawierające:

procesor zdolny do:

wygenerowania jednego lub więcej kluczy szyfrowania (140, 150) pochodzących z pierwszej sekwencji genomowej (650);

zaszyfrowania wiadomości (100) przy użyciu jednego lub każdego klucza szyfrowania (140, 150) w celu utworzenia zaszyfrowanej wiadomości;

wysłania zaszyfrowanej wiadomości do jednego lub większej liczby urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania (140, 150) pochodzących z drugiej sekwencji genomowej;

oraz

otrzymania potwierdzenia, czy odszyfrowanie zaszyfrowanej wiadomości powiodło się przez jedno lub więcej zdalnych urządzeń;

przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej (650).

15. System do lokalizowania jednego lub więcej członków sieci rodzinnej, obejmujący:

procesor zdolny do:

wygenerowania jednego lub więcej kluczy szyfrowania (140, 150) pochodzących z pierwszej sekwencji genomowej (650);

zaszyfrowania wiadomości (100) przy użyciu jednego lub każdego klucza szyfrowania (140, 150) w celu utworzenia zaszyfrowanej wiadomości;

wysłania zaszyfrowanej wiadomości do jednego lub większej liczby urządzeń, przy czym odszyfrowanie zaszyfrowanej wiadomości na jednym lub większej liczbie urządzeń wykorzystuje jeden lub większą liczbę kluczy szyfrowania (140, 150) pochodzących z drugiej sekwencji genomowej;

otrzymania potwierdzenia, czy odszyfrowanie zaszyfrowanej wiadomości powiodło się przez jedno lub więcej zdalnych urządzeń;

przy czym wysłanie zaszyfrowanej wiadomości obejmuje wygenerowanie adresu genetycznego z pierwszej sekwencji genomowej (650).

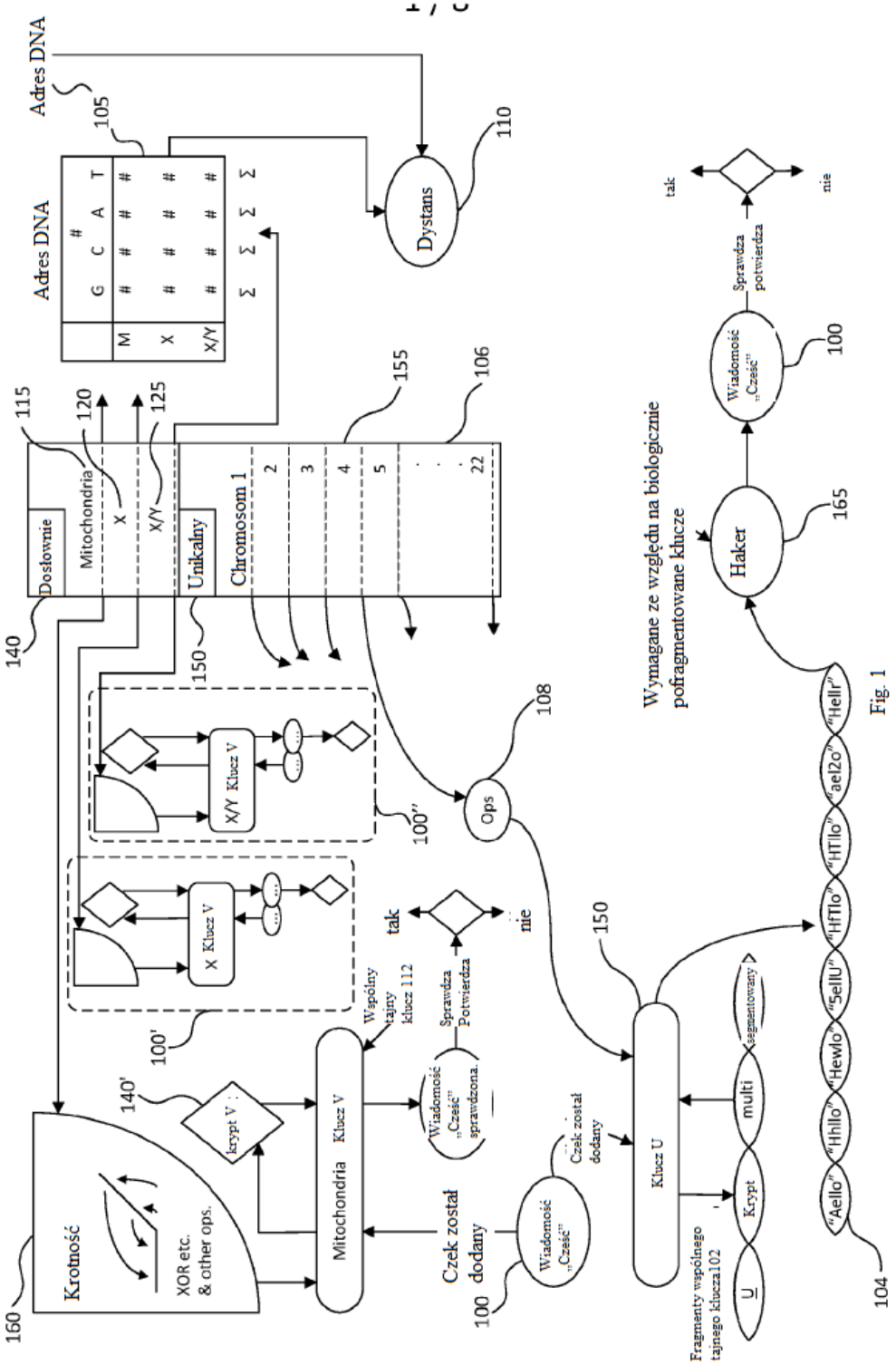
Adres DNA
105
Adres DNA
G C A T
M # # # #
X # # # #
X/Y # # # #
Σ Σ Σ Σ
Dystans
110
115
120
125
155
106
Dosłownie
Mitochondria
X
X/Y
Unikalny
Chromosom 1
2
3
4
5
22
140
150
100''
X/Y Klucz V
X Klucz V
100'
140'
krypt V
Mitochondria Klucz V
Wspólny tajny klucz 112
Sprawdza Potwierdza
tak
nie
Wiadomość „Cześć" sprawdzona
Czek został dodany
100
Wiadomość „Cześć"
Czek został dodany
160
Krotność
XOR etc. & other ops.
108
Ops
150
Klucz U
Fragmenty wspólnego tajnego klucza102
U
Krypt
multi
segmentowany
"Aello"
"Hhllo"
"Hewlo"
"5ellU"
"HfTlo"
"HTllo"
"ael2o"
"Hellr"
104
Wymagane ze względu na biologicznie pofragmentowane klucze
Haker
165
Wiadomość „Cześć"
100
Sprawdza potwierdza
tak
nie

Fig. 1

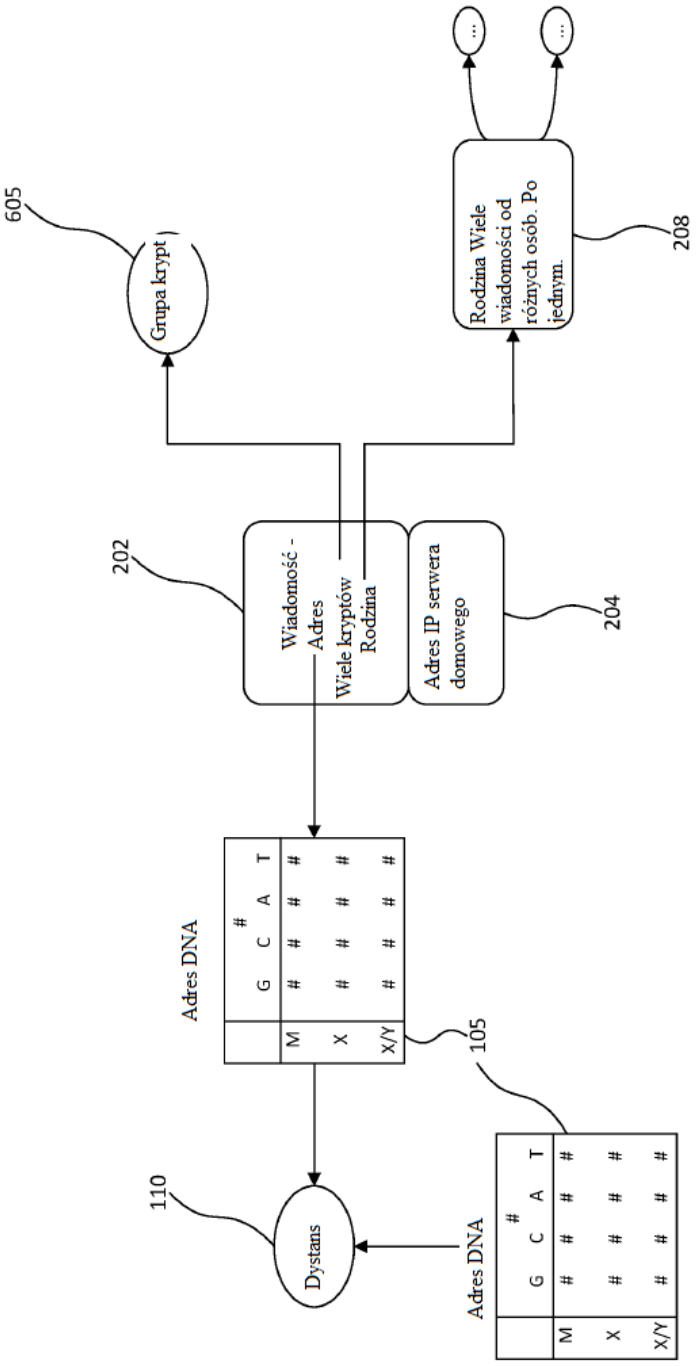
605
Grupa krypt)
208
Rodzina Wiele wiadomości od różnych osób. Po jednym.
...
...
202
Wiadomość - Adres
Wiele kryptów
Rodzina
Adres IP serwera domowego
204
Adres DNA
105
110
Dystans
Adres DNA

Fig. 2

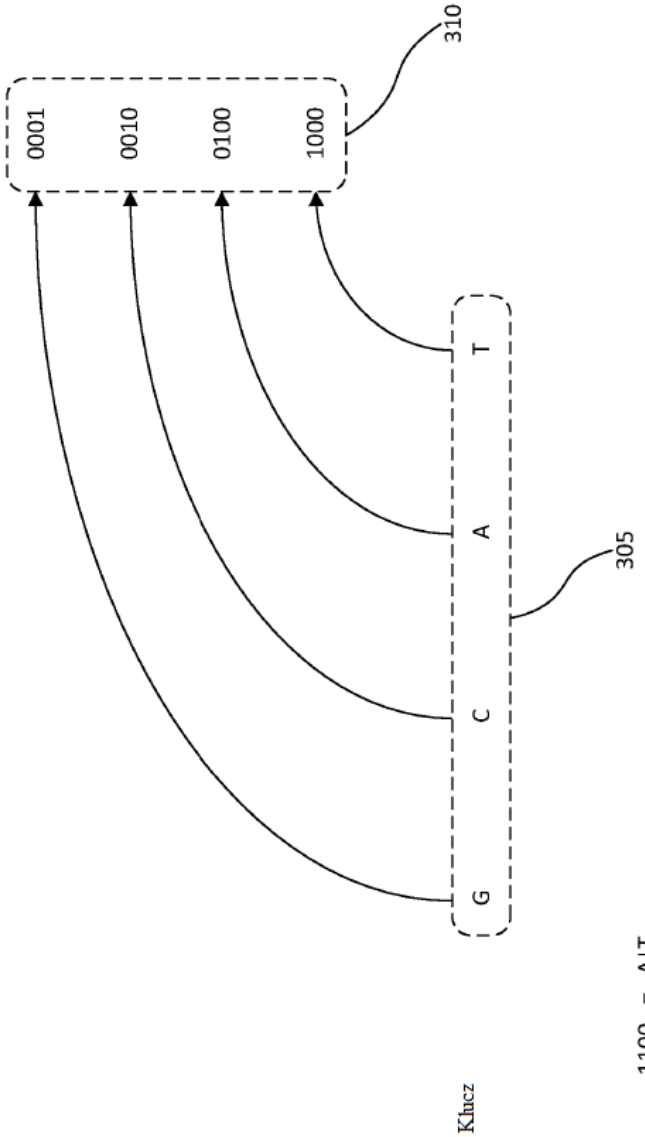
310
0001
0010
0100
1000
305
G
C
A
T
Klucz
1100 = A|T
Etc.
0 → f
Dla par zasad

Fig. 3

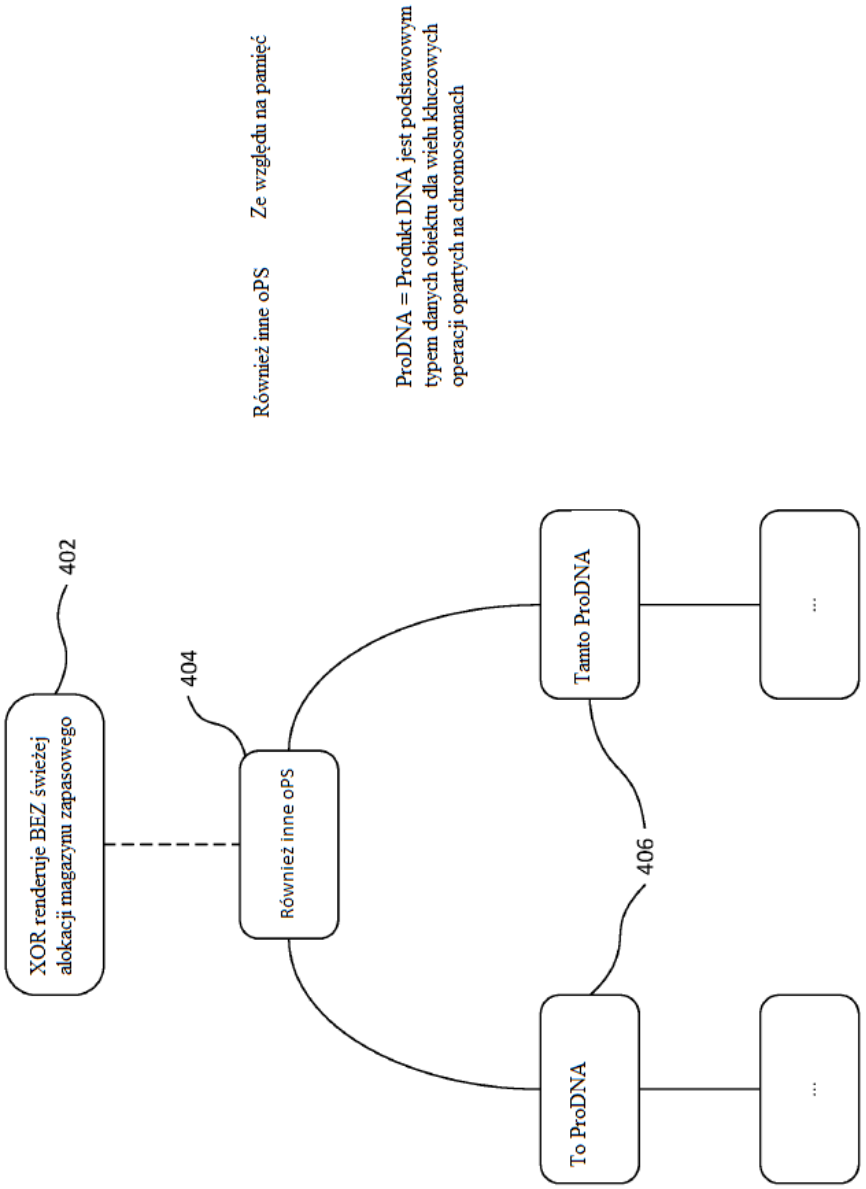
Również inne oPS
Ze względu na pamięć
ProDNA = Produkt DNA jest podstawowym typem danych obiektu dla wielu kluczowych operacji opartych na chromosomach
402
XOR renderuje BEZ świeżej alokacji magazynu zapasowego
404
Również inne oPS
Tamto ProDNA
...
406
To ProDNA
...

Fig. 4

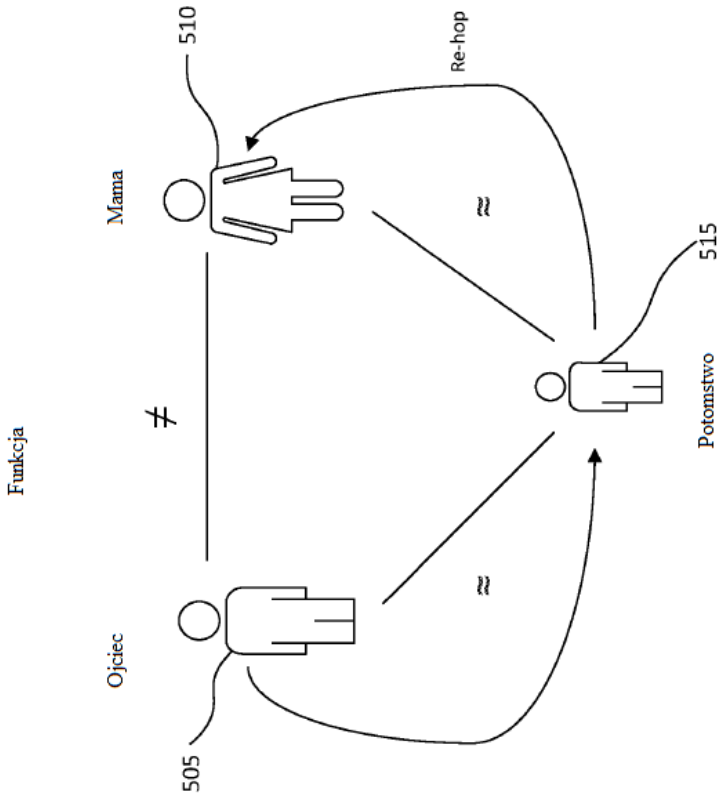

Fig. 5

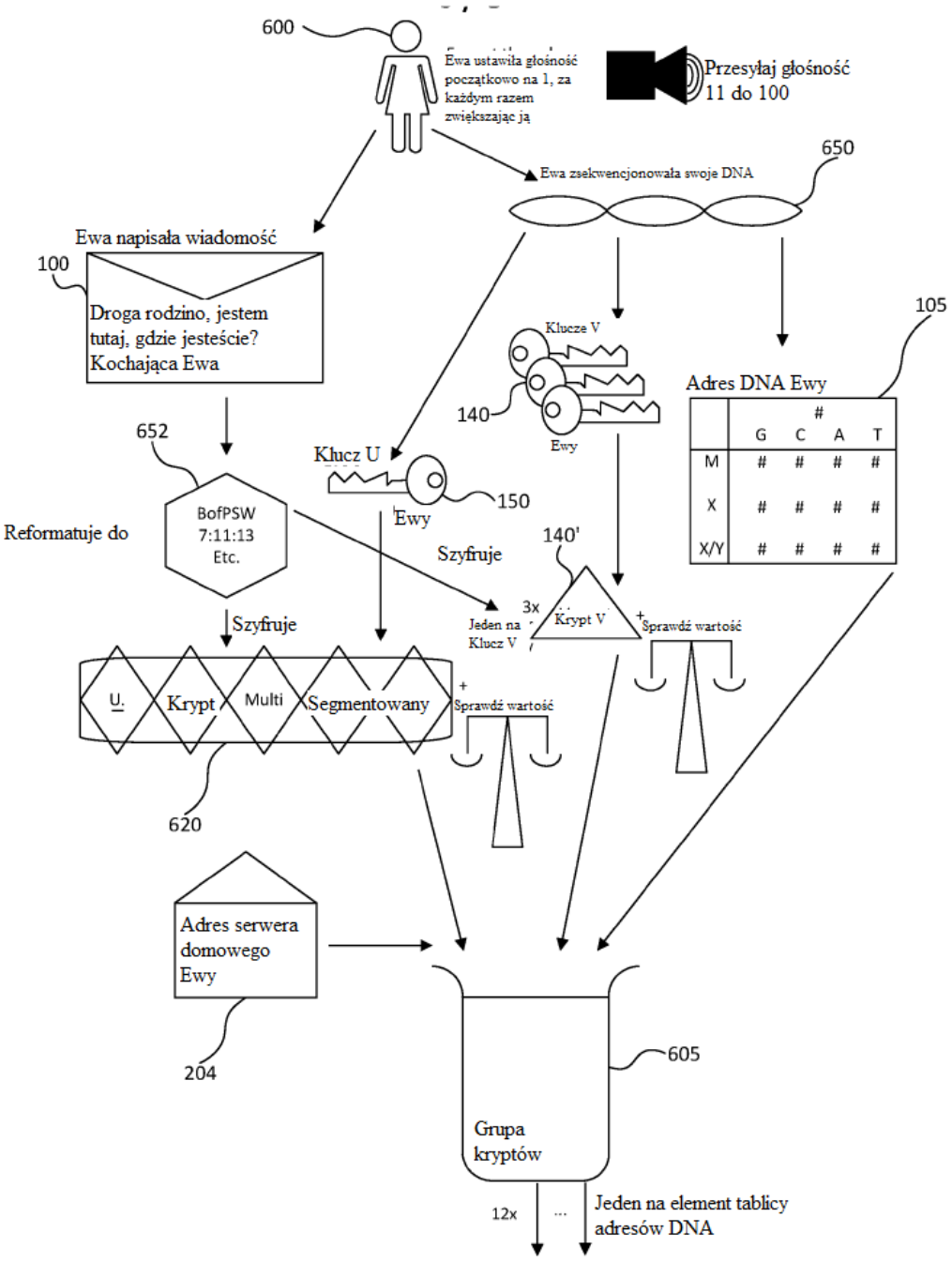
600
Ewa ustawiła głośność początkowo na 1, za każdym razem zwiększając ją
Przesyłaj głośność 11 do 100
650
Ewa zsekwencjonowała swoje DNA
Ewa napisała wiadomość
100
Droga rodzino, jestem tutaj, gdzie jesteście? Kochająca Ewa
105
Klucze V
140
Ewy
Adres DNA Ewy
G C A T
M # # # #
X # # # #
X/Y # # # #
652
Klucz U
150
Ewy
BofPSW 7:11:13 Etc.
Reformatuje do
Szyfruje
140'
3x
Jeden na Klucz V
Krypt V
+Sprawdź wartość
Szyfruje
U.
Krypt
Multi
Segmentowany
+Sprawdź wartość
620
Adres serwera domowego Ewy
204
605
Grupa kryptów
12x
...
Jeden na element tablicy adresów DNA

Fig. 6a

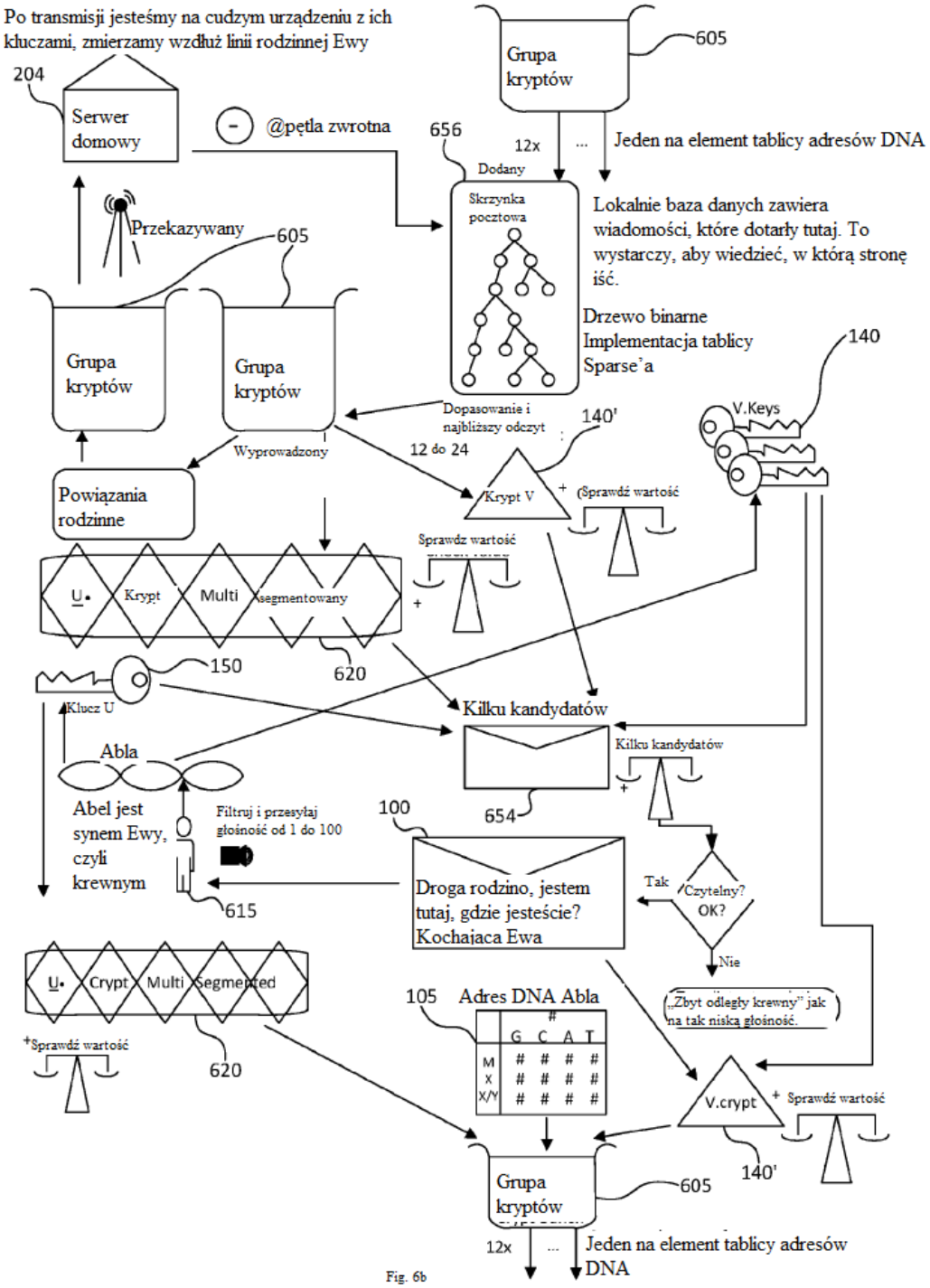
Po transmisji jesteśmy na cudzym urządzeniu z ich kluczami, zmierzamy wzdłuż linii rodzinnej Ewy
204
Serwer domowy
@pętla zwrotna
Grupa kryptów
605
656
12x
...
Jeden na element tablicy adresów DNA
Dodany
Skrzynka pocztowa
Lokalnie baza danych zawiera wiadomości, które dotarły tutaj. To wystarczy, aby wiedzieć, w którą stronę iść.
Drzewo binarne Implementacja tablicy Sparse'a
Przekazywany
605
Grupa kryptów
Grupa kryptów
Dopasowanie i najbliższy odczyt
140'
140
V.Keys
Wyprowadzony
12 do 24
Krypt V
+
Sprawdź wartość
Powiązania rodzinne
Sprawdź wartość
U•
Krypt
Multi
segmentowany
620
150
Klucz U
Kilku kandydatów
Kilku kandydatów
Abla
654
100
Abel jest synem Ewy, czyli krewnym
Filtruj i przesyłaj głośność od 1 do 100
615
Droga rodzino, jestem tutaj, gdzie jesteście? Kochajaca Ewa
Tak
Czytelny? OK?
Nie
„Zbyt odległy krewny" jak na tak niską głośność.
U•
Crypt
Multi
Segmented
+Sprawdź wartość
620
105
Adres DNA Abla
#
G C A T
M
X
X/Y
V.crypt
+ Sprawdź wartość
140'
Grupa kryptów
605
12x
...
Jeden na element tablicy adresów DNA

Fig. 6b

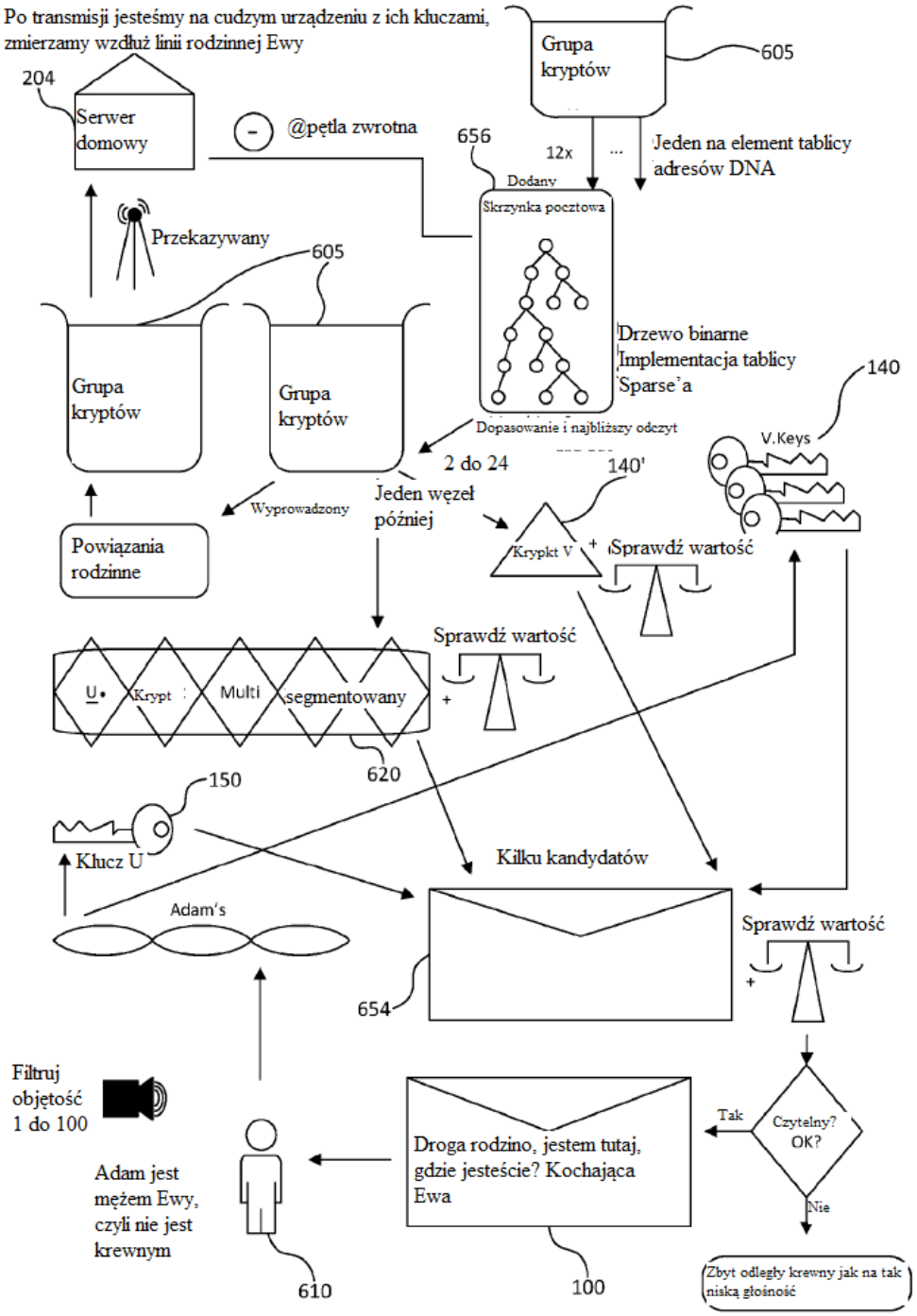
Po transmisji jesteśmy na cudzym urządzeniu z ich kluczami, zmierzamy wzdłuż linii rodzinnej Ewy
204
Serwer domowy
@pętla zwrotna
Grupa kryptów
605
656
12x
Jeden na element tablicy adresów DNA
Dodany
Skrzynka pocztowa
Przekazywany
Drzewo binarne Implementacja tablicy Sparse'a
Grupa kryptów
Grupa kryptów
Dopasowanie i najbliższy odczyt
140
V.Keys
2 do 24
140'
Jeden węzeł później
Wyprowadzony
Powiązania rodzinne
Krypkt V
Sprawdź wartość
Sprawdź wartość
U
Krypt
Multi
segmentowany
620
150
Klucz U
Kilku kandydatów
Adam's
Sprawdź wartość
654
Filtruj objętość 1 do 100
Adam jest mężem Ewy, czyli nie jest krewnym
610
Droga rodzino, jestem tutaj, gdzie jesteście? Kochająca Ewa
100
Tak
Czytelny? OK?
Nie
Zbyt odległy krewny jak na tak niską głośność

Fig. 6c

Ta lista odnośników cytowanych przez zgłaszającego podana jest wyłącznie dla wygody czytelnika. Nie stanowi ona części europejskiego dokumentu patentowego. Choć zestawianie tych odnośników przeprowadzono z wielką starannością, nie można wykluczyć błędów lub braków i EPO zrzeka się wszelkiej odpowiedzialności w tym zakresie.

Dokumenty patentowe cytowane w opisie

- US 2015112884 A1 [0004]
- US 2014289536 A1 [0005]