



(21) 申請案號：110112537

(22) 申請日：中華民國 110 (2021) 年 04 月 07 日

(51) Int. Cl. : **G16H30/20 (2018.01)****G06F17/18 (2006.01)****G06F12/0802(2016.01)****G16B15/10 (2019.01)**

(30) 優先權：2020/04/07 美國

63/006,668

(71) 申請人：美商宜曼達公司 (美國) ILLUMINA, INC. (US)

美國

(72) 發明人：魯勒 麥可 RUEHLE, MICHAEL (US)

(74) 代理人：李世章；彭國洋

(56) 參考文獻：

CN 108334750A

CN 109935274A

CN 110875084A

US 2017/0364666A1

US 2018/0137387A1

US 2020/0104464A1

審查人員：施易昉

申請專利範圍項數：20 項 圖式數：5 共 59 頁

(54) 名稱

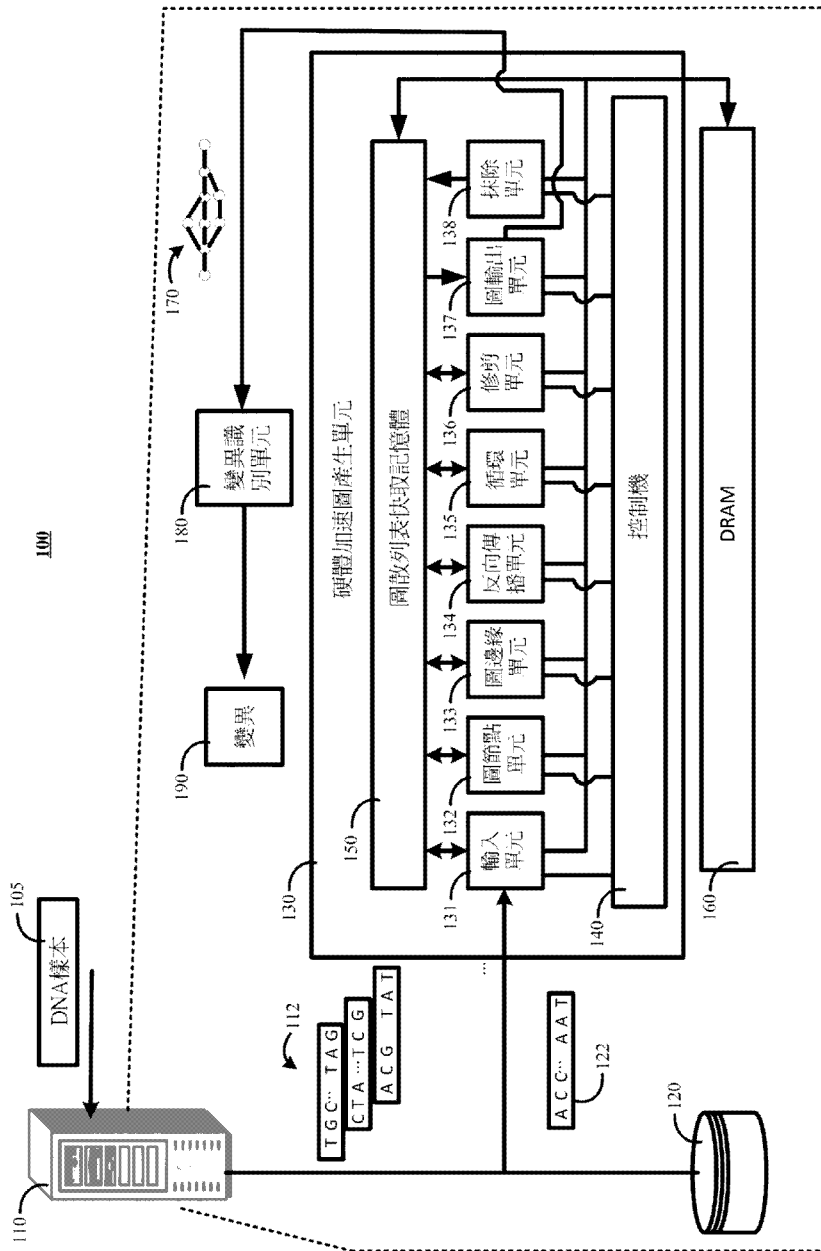
硬體加速的 K-MER 圖產生

(57) 摘要

用於使用一可程式化邏輯裝置硬體加速產生一 K-mer 圖的方法、系統、及設備。在一個態樣中，一種方法包括下列動作：獲得一第一組核酸序列；使用所獲得之該第一組核酸序列及使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元產生一 K-mer 圖；及在由各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該 K-mer 圖的圖描述資料。

Methods, systems, and apparatus for hardware-accelerated generation of a K-mer graph using a programmable logic device. In one aspect, a method includes actions of obtaining a first set of nucleic acid sequences, generating a K-mer graph using the obtained first set of nucleic acid sequences and using a plurality of non-pipelined hardware logic units of a programmable logic device, and periodically updating, with a control machine, graph description data for the K-mer graph after performance of the one or more operations by each hardware logic unit.

指定代表圖：



【圖 1】

符號簡單說明：

- 100:系統
- 105:生物樣本/DNA 樣本
- 110:核酸定序儀
- 112:片段
- 120:參考序列資料庫/參考基因體資料庫
- 122:參考基因體/參考基因體資料庫
- 130:硬體加速圖產生單元
- 131:輸入單元/硬體邏輯單元
- 132:圖節點單元/圖單元/硬體邏輯單元
- 133:圖邊緣單元/硬體邏輯單元
- 134:反向傳播單元/硬體邏輯單元
- 135:循環單元/硬體邏輯單元
- 136:修剪單元/硬體邏輯單元
- 137:圖輸出單元/硬體邏輯單元
- 138:抹除單元/硬體邏輯單元
- 140:控制機
- 150:快取記憶體
- 160:DRAM
- 170:K-mer 圖
- 180:變異識別單元
- 190:變異



I867202

【發明摘要】

【中文發明名稱】 硬體加速的K-MER圖產生

【英文發明名稱】 HARDWARE ACCELERATED K-MER GRAPH
GENERATION

【中文】

用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖的方法、系統、及設備。在一個態樣中，一種方法包括下列動作：獲得一第一組核酸序列；使用所獲得之該第一組核酸序列及使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元產生一K-mer圖；及在由各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該K-mer圖的圖描述資料。

【英文】

Methods, systems, and apparatus for hardware-accelerated generation of a K-mer graph using a programmable logic device. In one aspect, a method includes actions of obtaining a first set of nucleic acid sequences, generating a K-mer graph using the obtained first set of nucleic acid sequences and using a plurality of non-pipelined hardware logic units of a programmable logic device, and periodically updating, with a control machine, graph description data for the K-mer graph after performance of the one or more operations by each hardware logic unit.

【指定代表圖】第(1)圖

【代表圖之符號簡單說明】

100:系統

- 105:生物樣本/DNA樣本
- 110:核酸定序儀
- 112:片段
- 120:參考序列資料庫/參考基因體資料庫
- 122:參考基因體/參考基因體資料庫
- 130:硬體加速圖產生單元
- 131:輸入單元/硬體邏輯單元
- 132:圖節點單元/圖單元/硬體邏輯單元
- 133:圖邊緣單元/硬體邏輯單元
- 134:反向傳播單元/硬體邏輯單元
- 135:循環單元/硬體邏輯單元
- 136:修剪單元/硬體邏輯單元
- 137:圖輸出單元/硬體邏輯單元
- 138:抹除單元/硬體邏輯單元
- 140:控制機
- 150:快取記憶體
- 160:DRAM
- 170:K-mer圖
- 180:變異識別單元
- 190:變異

【特徵化學式】

無

【發明說明書】

【中文發明名稱】 硬體加速的K-MER圖產生

【英文發明名稱】 HARDWARE ACCELERATED K-MER GRAPH

GENERATION

【技術領域】

本揭示案係關於硬體加速的K-MER圖產生。

【先前技術】

【0001】 K-mer圖可用以表示複數個定序片段。

【發明內容】

【0002】 根據本揭露的一個新穎態樣，揭示一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之方法。在一個態樣中，該方法可包括下列動作：獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一K-mer圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該K-mer圖的各節點表示一K-mer，該K-mer圖的各邊緣表示一對K-mer之間的一鏈接，且該K-mer圖之各邊緣的各權重表示藉由一對K-mer表示的一K-mer序列的一發生數目；及在該K-mer圖的產生期間：在由用以產生該K-mer圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控

制機週期地更新用於該K-mer圖的圖描述資料，其中該圖描述資料表示(i)一K-mer圖識別符，及(ii)K-mer圖狀態資訊，其中該控制機藉由在該K-mer圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【0003】 其他版本包括已組態以執行藉由一硬體加速圖產生單元的硬體邏輯電路定義的前述方法的該等動作的對應系統及設備。

【0004】 此等及其他版本可選地包括下列特徵的一或多者。例如，在一些實施方案中，該複數個硬體邏輯單元的各硬體邏輯單元的該輸出係經由一散列表快取記憶體儲存。

【0005】 在一些實施方案中，該控制機係使用該可程式化邏輯裝置的一硬體邏輯單元實施。

【0006】 在一些實施方案中，該控制機係使用一或多個CPU或GPU執行軟體指令以實現該控制機的功能性而實施。

【0007】 在一些實施方案中，操作可進一步包括將所產生之該K-mer圖提供至一變異識別(variant calling)單元，其中該變異識別單元處理該K-mer圖以判定該複數個片段的一或多者與該參考序列之間的候選變異。

【0008】 在一些實施方案中，可藉由一或多個CPU或GPU執行軟體指令以實現該變異識別單元的一或多個功能。

【0009】 在一些實施方案中，可程式化邏輯裝置係用以加速該變異識別單元的一或多個功能。

【0010】 在一些實施方案中，該圖描述資料進一步包括(iii)表示該複數個硬體邏輯單元之在該K-mer圖上或在與該K-mer圖識別符關聯之該堆積的核酸序列上執行硬體邏輯的一最末硬體邏輯單元的資料。

【0011】 根據本揭露的另一新穎態樣，揭示一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之系統。在一個態樣中，該系統可包括一硬體加速圖產生單元，該硬體加速圖產生單元包括已配置以執行操作的硬體數位邏輯電路。在一些實施方案中，該操作可包含：獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一K-mer圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該K-mer圖的各節點表示一K-mer，該K-mer圖的各邊緣表示一對K-mer之間的一鏈接，且該K-mer圖之各邊緣的各權重表示藉由一對K-mer表示的一K-mer序列的一發生數目；及在該K-mer圖的產生期間：在由用以產生該K-mer圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該K-mer圖的圖描述資料，其中該圖描述資料表示(i)一K-mer圖識別符，及(ii)K-mer圖狀態資訊，其中該控制機藉由在該K-mer圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【0012】 其他版本包括用於執行前述操作的對應方法及設備。

【0013】 此等及其他版本可選地包括下列特徵的一或多者。例如，在一些實施方案中，該複數個硬體邏輯單元的各硬體邏輯單元的該輸出係經由一散列表快取記憶體儲存。

【0014】 在一些實施方案中，該系統可進一步包括一或多個電腦，及一或多個記憶體裝置，該一或多個記憶體裝置儲存指令，該等指令當由該一或多個電腦執行時，導致該一或多個電腦執行一變異識別單元的第二操作。在一些實施方案中，該變異識別單元的該等第二操作可包括藉由該變異識別單元獲得所產生之該K-mer圖，及基於該變異識別單元處理所產生之該K-mer圖而識別一或多個候選變異，其中一候選變異係在片段之該堆積中的一或多個片段的一鹼基識別與在一參考基因體的一特定位置的該參考基因體的一核苷酸之間的一差異。

【0015】 在一些實施方案中，操作可進一步包括藉由一變異識別單元獲得所產生之該K-mer圖，及基於該變異識別單元處理所產生之該K-mer圖而識別一或多個候選變異，其中一候選變異係在片段之該堆積中的一或多個片段的一鹼基識別與在一參考基因體的一特定位置的該參考基因體的一核苷酸之間的一差異。

【0016】 在一些實施方案中，該圖描述資料可進一步包括(iii)表示該複數個硬體邏輯單元之在該K-mer圖上或在與該K-mer圖識別符關聯之該堆積的核酸序列上執行硬體邏輯的一最末硬體邏輯單元的資料。

【0017】 根據本揭露的另一新穎態樣，揭示一種硬體加速圖產生單元。在一個態樣中，該硬體加速圖產生單元可包括已配置以執行操作的硬體數位邏輯電路。在一些實施方案中，該等操作可包括：獲得一第一組核酸序列，其中

該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一K-mer圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該K-mer圖的各節點表示一K-mer，該K-mer圖的各邊緣表示一對K-mer之間的一鏈接，且該K-mer圖之各邊緣的各權重表示藉由一對K-mer表示的一K-mer序列的一發生數目；及在該K-mer圖的產生期間：在由用以產生該K-mer圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該K-mer圖的圖描述資料，其中該圖描述資料表示(i)一K-mer圖識別符，及(ii)K-mer圖狀態資訊，其中該控制機藉由在該K-mer圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【0018】 根據本揭露的另一新穎態樣，揭示一種用於在一可程式化邏輯裝置中硬體加速產生一K-mer圖之方法。在一個態樣中，該方法可包括下列動作：獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；對於該第一組核酸序列的各特定核酸序列：針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各K-mer的一圖節點的資料；藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各K-mer的一圖節點的產生；藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及對於一或多對所產生之該等圖節點：藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或

多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各K-mer的該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一K-mer圖。

【0019】 其他版本包括已組態以執行藉由一硬體加速圖產生單元的硬體邏輯電路定義的前述方法的該等動作的對應系統及設備。

【0020】 此等及其他版本可選地包括下列特徵的一或多者。例如，在一些實施方案中，該方法可進一步包括評估該K-mer圖以檢查圖循環的存在。在此類實施方案中，若在該評估期間偵測到一圖循環，則該程序可包括終止該K-mer圖的產生。替代地，若在該評估期間未偵測到一圖循環：該方法可包括從該散列表快取記憶體獲得描述該K-mer圖的該結構的資料，並將描述該K-mer圖的該結構的所獲得之該資料提供至一變異識別模組。

【0021】 在一些實施方案中，方法進一步可進一步包括藉由該控制機且在可由該控制機存取的一記憶體單元中週期地儲存用於該K-mer圖的一實例的圖描述資料，其中該圖描述資料表示(i)一K-mer圖識別符，及(ii)K-mer圖狀態資訊。

【0022】 在一些實施方案中，該第一硬體邏輯單元可進一步經組態以：判定該特定核酸序列的該等特定K-mer的一或多者是否與該特定核酸序列的另一K-mer匹配，並基於該特定核酸序列的該一或多個特定K-mer與該特定核酸序列的另一K-mer匹配的一判定，儲存將該一或多個特定K-mer標記成非獨特K-mer的資料。

【0023】 在一些實施方案中，該第二硬體邏輯進一步經組態以：將一邊緣權重指派給該K-mer圖的各邊緣。

【0024】 在一些實施方案中，該方法可進一步包括指示該可程式化邏輯裝置的一第三硬體邏輯單元執行經組態以：從該散列表快取記憶體獲得表示該K-mer圖的資料，並提供表示該K-mer圖的所獲得之該資料至一變異識別單元的硬體邏輯。

【0025】 在一些實施方案中，該方法可進一步包括指示該可程式化邏輯裝置的一第三硬體邏輯單元執行經組態以：從該散列表快取記憶體選擇性地刪除表示該K-mer圖之圖節點的資料及表示該K-mer圖之圖邊緣的資料的硬體邏輯。

【0026】 在一些實施方案中，該控制機係使用該可程式化邏輯裝置的一第三硬體邏輯單元實施。

【0027】 在一些實施方案中，該散列表快取記憶體係使用該可程式化邏輯裝置的一第三硬體邏輯單元實施。

【0028】 在一些實施方案中，該控制機係藉由使用執行軟體指令以實現該控制機的功能性的一或多個CPU或GPU而實施。

【0029】 在一些實施方案中，該圖描述資料進一步包括(iii)表示該複數個硬體邏輯單元之在該K-mer圖上或在與該K-mer圖識別符關聯之該堆積的核酸序列上執行硬體邏輯的一最末硬體邏輯單元的資料。

【0030】 根據本揭露的另一新穎態樣，揭示一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之系統。該系統可包括一硬體加速圖產生單元，該硬體加速圖產生單元包括已配置以執行操作的硬體數位邏輯電路。在一個態樣中，該等操作可包括：獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部

分；對於該第一組核酸序列的各特定核酸序列：針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各K-mer的一圖節點的資料；藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各K-mer的一圖節點的產生；藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及對於一或多對所產生之該等圖節點：藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各K-mer的該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一K-mer圖。

【0031】 其他版本包括用於執行前述操作的對應方法及設備。

【0032】 此等及其他版本可可選地包括下列特徵的一或多者。例如，在一些實施方案中，該等操作可進一步包括藉由該控制機且在可由該控制機存取的一記憶體單元中週期地儲存用於該K-mer圖的一實例的圖描述資料，其中該圖描述資料表示(i)一K-mer圖識別符，及(ii)K-mer圖狀態資訊。

【0033】 在一些實施方案中，該第一硬體邏輯單元進一步經組態以：判定該特定核酸序列的該等特定K-mer的一或多者是否與該特定核酸序列的另一K-mer匹配，並基於該特定核酸序列的該一或多個特定K-mer與該特定核酸序列的另一K-mer匹配的一判定，儲存將該一或多個特定K-mer標記成非獨特K-mer的資料。在此類實施方案中，該第二硬體邏輯進一步經組態以：將一邊緣權重指派給該K-mer圖的各邊緣。

【0034】 根據本揭露的另一新穎態樣，揭示一種硬體加速圖產生單元。

該硬體加速圖產生單元可包括已配置以執行操作的硬體數位邏輯電路。在一個態樣中，該等操作可包括：獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；對於該第一組核酸序列的各特定核酸序列：針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各K-mer的一圖節點的資料；藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各K-mer的一圖節點的產生；藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及對於一或多對所產生之該等圖節點：藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各K-mer的該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一K-mer圖。

【0035】 本揭露的此等及其他態樣在下文的實施方式中參照隨附圖式更詳細地討論。

【圖式簡單說明】

【0036】

〔圖1〕係用於硬體加速產生K-mer圖之系統的實例。

〔圖2〕係用於硬體加速產生K-mer圖之程序的實例的流程圖。

〔圖3〕係用於硬體加速產生K-mer圖之程序的另一實例的流程圖。

〔圖4〕係K-mer圖的實例。

〔圖5〕係可用於硬體加速K-mer圖之系統組件的實例的方塊圖。

【實施方式】

【0037】 本申請案主張於2020年4月7日申請之美國臨時專利申請案第63/006,668號之優先權，其全文內容以引用方式併入本文中。

【0038】 本揭露係關於K-mer圖的硬體加速產生。使用硬體電路產生K-mer圖顯著地降低產生K-mer圖所需的時間量並將計算密集的K-mer圖產生程序從軟體處理器卸載至積體電路（諸如現場可程式化閘陣列或ASIC）的硬體邏輯。此釋放可用以執行其他基因體資料處理任務之軟體處理器的軟體資源。

【0039】 K-mer圖的硬體加速產生可使用控制機達成，該控制機經組態以管理由複數個非管線化硬體邏輯單元執行之操作的工作流程。具體而言，控制機可使用複數個非管線化硬體邏輯單元抽象地達成高階管線化功能性。控制機可藉由儲存及更新圖描述資料來達成此功能性，該圖描述資料包括(i)識別K-mer圖之實例的K-mer圖識別符及(ii)K-mer圖狀態資訊。K-mer圖狀態資訊可包括，例如，指示在用於K-mer圖之特定實例的原始圖資料上操作的最末硬體邏輯單元的資料、指示最末硬體邏輯單元是否中止操作的資料、指示K-mer長度的資料、指示K-mer節點列表的資料、指示可用以識別快取記憶體中之K-mer節點的指標器之列表的資料、K-mer節點列表的長度、指示非獨特K-mer之列表的資料、或其任何子集或組合。控制機藉由呼叫待在原始圖資料上執行操作的特定硬體邏輯單元，並將經更新的圖描述資料組提供至被呼叫的硬體邏輯單元來管理K-mer圖產生。

【0040】 圖描述資料的此儲存及更新使控制機能以使非管線化硬體邏輯單元之各者能在對應於不同K-mer圖之資料上操作的方式管理非管線化硬體邏輯單元的平行處理。因此，除了使用硬體邏輯取代軟體指令的執行以產生K-mer圖而達成增加速度的利益外，本揭露藉由在相同時間使用由控制機管理的不同硬體邏輯單元產生不同K-mer圖的區段來增加生產量而達成操作的進一步加速。

【0041】 圖1係用於硬體加速產生K-mer圖之系統100的實例。在一些實施方案中，系統100可包括核酸定序儀110、參考序列資料庫120、硬體加速圖產生單元130、複數個硬體邏輯單元131、132、133、134、135、136、137、138、控制機140、圖散列表快取記憶體150、DRAM 160、及變異識別單元180。在一些實施方案中，硬體加速圖產生單元130可使用可程式化邏輯電路（諸如現場可程式化閘陣列(field programmable gate array, FPGA)）實施。在其他實施方案中，硬體加速圖產生單元130可使用特定應用積體電路(Application Specification Integrated Circuit, ASIC)實施。在任一場景中，相關於硬體加速圖產生單元130及實施於其上之組件之各者描述的功能性係在不執行軟體指令實現功能性的情況下使用經配置以實現本文描述之功能性的硬體邏輯電路來實施。

【0042】 用語「單元(unit)」在本說明書中係用以描述用以執行指定功能的軟體模組、硬體模組、或二者之組合。在本文中描述之特定「單元」是否係硬體、軟體、或二者之組合的判定可基於其使用的上下文作出。例如，常駐在使用FPGA或ASIC實施之硬體加速圖產生單元130中的「輸入單元」131、「圖節點單元」132、「圖邊緣單元」133、或類似者係其功能性係藉由已配置以實

現相關於該特定「單元」於本文描述之功能性的固線式數位邏輯閘或固線式數位邏輯區塊實現的硬體單元。藉由另一實例，非使用圖1之硬體加速圖產生單元130實施的「變異識別單元」180係其功能性係藉由執行定義「變異識別單元」180之功能性的軟體指令的一或多個電腦來實現的軟體模組。藉由另一實例，電腦或處理單元可係藉由處理軟體指令實現功能性的硬體裝置，且因此電腦或處理單元的功能性係硬體及軟體的組合。

【0043】 雖然因為「控制機」在圖1中係描繪成在硬體加速圖產生單元130中實施，圖1的一或多個組件（諸如「控制機」140）的實例在本文中提供為硬體實施方案，但本揭露不受限於此類實例。替代地，可採用其他實施方案，其中「控制機」140係以軟體實施為軟體模組或硬體及軟體的組合，搭配執行軟體指令的電腦或處理單元以實現本文描述之「控制機」140之功能性。同樣地，本揭露可有將相關於圖1描述成軟體的某些組件（諸如「變異識別單元」180）實施為硬體實施方案的實施方案。

【0044】 核酸定序儀110係經組態以執行初步分析的裝置。初步分析可包括由核酸定序儀110接收生物樣本105（諸如血液樣本、組織樣本、唾液、或核酸），並由核酸定序儀110產生輸出資料（諸如各自表示所接收生物樣本之核酸序列的核苷酸順序的一或多個片段112）。在一些實施方案中，藉由核酸定序儀110的定序可在多個片段循環中執行，其中第一片段循環產生包括從核酸序列碎片的第一端表示核苷酸順序的鹼基識別字串的一或多個第一片段，且第二片段循環產生包括從該等核酸序列碎片的一者的另一端表示核苷酸順序的鹼基識別字串的一或多個各別第二片段。在一些實施方案中，片段可使用複製擴

增產生。在圖1的實例中，一或多個片段112可包括特定參考基因體位置之片段的堆積，其中參考基因體位置包含多個循序參考基因體位置。

【0045】 因此，各片段係表示有機體（諸如動物、昆蟲、植物、或類似者）之核酸基因體的一部分的資料。假設大約600個鹼基識別的短核酸序列碎片，第一片段可表示核酸序列碎片之第一端的150個有序核苷酸，且第二片段可表示該核酸序列碎片之另一端的150個有序核苷酸。然而，此等數目僅係實例，且任何核酸定序儀110可經組態以產生可使用任何定序方法藉由如本文描述之硬體加速圖產生單元於其上操作的片段。此類片段可係與本文提及之長度不同的長度的。例如，在一些實施方案中，本揭露可用以產生由具有至多1000個或更多個核苷酸的核酸序列碎片產生之片段的硬體加速K-mer圖，其中各片段具有，例如，來自各碎片之該端的50個鹼基識別、75個鹼基識別、150個鹼基識別、200個鹼基識別、300個鹼基識別、500個鹼基識別、或更多。各鹼基識別可對應於一核苷酸。本揭露亦可用以產生長片段的硬體加速K-mer圖。因此，硬體加速圖產生單元130可用以產生藉由任何類型的核酸定序儀以任何方式產生之任何片段的K-mer圖。

【0046】 在一些實施方案中，生物樣本105可包括DNA樣本，且核酸定序儀110可包括DNA定序儀。在此類實施方案中，由核酸定序儀產生之片段中的經定序核苷酸順序可包括鳥嘌呤(G)、胞嘧啶(C)、腺嘌呤(A)、及胸腺嘧啶(T)的一或多者。在一些實施方案中，核酸定序儀110可用以定序RNA樣本。在一些實施方案中，此可使用RNA-定序協定發生。舉實例而言，RNA樣本可使用反轉錄預處理，以使用反轉錄酶形成互補DNA (cDNA)。在其他實施方案中，核酸定序儀110可包括RNA定序儀，且生物樣本可包括RNA樣本。因此，雖然

圖1的實例描述產生包含由DNA定序儀基於DNA樣本產生之G、C、A、及T之片段的核酸定序儀，本揭露未如此受限。替代地，其他實施方案可處理由RNA定序儀基於RNA樣本產生的包含C、G、A、及U的片段。在一些實施方案中，由核酸定序儀產生的DNA片段或RNA片段可包括鹼基識別N，其中N指示由核酸定序儀產生的未知鹼基識別。

【0047】 在一些實施方案中，核酸定序儀110可包括經組態以通過大量平行定序技術的使用實現超高生產量、可縮放性、及速度的方式產生給定樣本的序列片段（諸如，片段112）的次一時代的定序儀(NGS)。NGS促成全基因體的快速定序、放大深度定序目標區域的能力、利用RNA定序(RNA-Seq)以發現新穎的RNA變異及剪接位、或量化mRNA以用於基因表現分析、表觀遺傳因子的分析（諸如全基因體DNA甲基化及DNA蛋白質交互作用）、癌症樣本的定序以研究罕見體細胞變異及腫瘤亞複製、及人類或環境中的微生物多樣性的研究。

【0048】 核酸定序儀110可從參考基因體資料庫122獲得參考基因體122。在一些實施方案中，僅獲得參考基因體122的一部分。所獲得的參考基因體122的部分可對應於片段112之堆積所映射並對準的參考基因體122的參考位置。參考基因體資料庫122可包括組織用於複數個不同參考基因體之儲存的資料儲存器。在一些實施方案中，選自參考基因體資料庫的特定參考基因體122可基於DNA樣本105的類型。在一些實施方案中，選自參考基因體資料庫120之所選參考基因體122的類型可基於來自核酸定序儀110之使用者的輸入而選擇。在此類實施方案中，使用者可，例如，選擇可由核酸定序儀110使用的參考基因體120識別符以從參考基因體資料庫120選擇特定參考基因體122。參考基因體122可包括，例如，組裝成特定物種之基因組的代表性實例的核酸序列。

【0049】 可將由核酸定序儀110產生之片段112的堆積及所獲得參考基因體122的組合提供為至硬體加速圖產生單元130的輸入。此等輸入可由硬體加速圖產生單元130之硬體邏輯單元131至138的一或多者處理以產生K-mer圖的實例。例如，硬體邏輯單元131至138的各硬體邏輯單元可經組態以針對包括成至硬體加速圖產生單元130的輸入之片段112的堆積的各片段執行其各別操作。

【0050】 在本文中將圖1的系統100描述成包括核酸定序儀。在一些實施方案中，諸如參照圖1所描述者，系統可包括定序儀110及硬體加速圖產生單元130，且系統100的其他組件可整合在核酸定序儀110內。然而，本揭露不限於整合在核酸定序儀110內。替代地，在一些實施方案中，硬體加速圖產生單元130可在整合或容納在電腦內的可程式化邏輯裝置或ASIC中實施，該電腦遠離核酸定序儀110並通訊地耦接至核酸定序儀110，諸如藉由使用一或多個有線或無線網路。類似地，資料庫120、變異識別單元180、或二者可在核酸定序儀之外實施。110。同樣地，系統100完全沒有包括核酸定序儀110的需要。替代地，在一些實施方案中，圖1的硬體加速圖產生單元130及其他組件可在不包括核酸定序儀110的電腦系統中實施。在此類實施方案中，硬體加速圖產生單元可經由網路從一或多個記憶體裝置或類似者的（多個）儲存位置獲得片段112的堆積、參考序列122、或二者。因此，系統100描繪本揭露的一實例，但不將本揭露限制在系統組件的任何一個具體組態。

【0051】 硬體加速圖產生單元130的一或多個硬體邏輯單元可包括輸入單元131、圖節點單元132、圖邊緣單元133、反向傳播單元134、循環單元135、修剪單元136、圖輸出單元137、及抹除單元138。在一些實施方案中，藉由硬體加速圖產生單元130產生K-mer圖可包括控制機140啟動及組態硬體邏輯單元

131至138之各者，以在儲存在快取記憶體150或DRAM 160中的一組資料上執行其等各別的硬體邏輯操作。在其他實施方案中，控制機140可僅啟動及組態硬體邏輯單元131至138的子集，以在儲存在快取記憶體150或DRAM 160中的一組原始圖資料上執行其等各別的硬體邏輯操作。

【0052】 舉實例而言，在一些實施方案中，硬體加速圖產生單元130可用以產生特殊化形式的德布魯因(De Bruijn)圖。可將此特殊化形式的德布魯因圖最佳化使得非獨特K-mer係使用圖中之各自具有單一邊緣的多個各別節點表示而非藉由具有多個邊緣的單一節點表示。此可部分地藉由使用圖單元132識別非獨特K-mer並標記非獨特K-mer用於進一步處理而達成。可將非獨特K-mer定義成在任何單一片段中發生至少二次或在參考序列中發生至少二次的K-mer序列。獨特K-mer不在相同片段中發生一次以上，但仍可能在多個片段中發生。

【0053】 然而，在其他實施方案中，德布魯因圖可在不在獨特或非獨特K-mer之間區分的情況下產生。因此，在一些實施方案中，不必實施可識別非獨特K-mer的圖節點單元132。在又另一實例中，反向傳播單元134不需要用以產生所有的K-mer圖。替代地，反向傳播單元134可限於可改善效能的實施方案。舉實例而言，當預期所產生之K-mer圖稍後變換成序列圖時，反向傳播單元134可用以改善邊緣權重的品質。

【0054】 相關於圖1之實例描述的K-mer圖產生實例顯示可由控制機140啟動及組態的各硬體邏輯單元131至138。在本說明中，通常將各硬體邏輯單元131至138描述成由控制機140啟動、由控制機140使用，例如，由控制機140儲存的圖描述資料組態、獲得原始圖資料、在所獲得原始圖資料或其他資料上執行一或多個特定處理操作、及接著更新原始圖資料、圖描述資料、或二者，本

揭露並未如此受限。然而，本揭露不限於此類實施方案。替代地，在一些實施方案中，各硬體邏輯單元131至138可經組態以執行其各別功能的多個實例。例如，圖節點單元132可經組態以接受至多3個不同的原始圖資料組且於其上同時執行其操作，循環硬體邏輯單元135可經組態以接受至多3個不同的原始圖資料組且於其上同時執行其操作，且PRU可經組態以接受至多2個不同的原始圖資料組且於其上同時執行其操作。各自對應於不同K-mer圖的不同組的數目或原始圖資料可由僅受限於系統100的可用硬體資源的特定硬體邏輯單元131至138接收及處理。例如，假設足夠位準的DRAM及FPGA邏輯單元可用於使用，可由硬體邏輯單元131至138接收及同時處理的不同的原始圖資料組的數目可大於3。同樣地，若此類資源非隨時可用或若不預期特定硬體邏輯單元將重度使用，一或多個硬體邏輯單元131至138可經組態以接收及同時處理較低數目的原始圖資料組。

【0055】 在又其他實施方案中，各能夠同時處理各對應於不同K-mer圖的不同的原始圖資料組的各硬體邏輯單元131至138不需要僅有一個實例。替代地，在一些實施方案中，各硬體加速圖產生單元130的多個實例可經組態以包括硬體邏輯單元131至138的一或多者的多個實例。在此類實例中，控制機140可經組態以監測各硬體邏輯單元131至138的狀態及可用性，並接著以橫跨各各別硬體邏輯單元負載平衡處理操作的方式啟動及組態各硬體邏輯單元。例如，在一些實施方案中，硬體加速圖產生單元130可經組態以具有各能夠接收至多3個不同的原始圖資料組並同時於其上執行其操作的圖節點單元132的3個實例、各能夠接收至多3個不同的原始圖資料組並同時於其上執行其操作的圖邊緣單元133的3個實例、各能夠接收至多2個不同的原始圖資料組並同時於其

等上執行其操作的2個反向傳播單元134、及各能夠接收至多2個不同的原始圖資料組並同時於其等上執行其操作的3個循環單元135。各硬體邏輯單元的啟動/停用、各硬體邏輯單元的組態、至各硬體邏輯單元的輸入、來自各硬體邏輯單元的輸出、及藉由各硬體邏輯單元更新圖描述資料係由控制機140管理及引導。

輸入單元

【0056】 輸入單元131可接收包括所產生的片段112之堆積及經選擇參考基因體122的輸入資料，該輸入資料在本文中可稱為原始圖資料。經選擇參考基因體122可包括參考基因體的一部分。原始圖資料可包括，例如，在K-mer圖之實例的產生期間由一或多個硬體邏輯單元131至138處理的資料。雖然原始圖資料包括，例如，經產生片段112及經選擇參考基因體122，原始圖資料亦可包括，例如，由圖節點單元132產生的K-mer節點、由圖邊緣單元133產生的邊緣、及類似者。輸入單元131可格式化用於儲存在DRAM 160的經產生片段112及經獲得參考基因體122。格式化經產生片段可包括，例如，編碼用於儲存於DRAM 160中的片段。在一些實施方案中，編碼片段可包括將對應於片段之核苷酸的各鹼基識別編碼成4位元的值。例如，A可編碼成0000、C可編碼成0001、G可編碼成0010、T可編碼成0011、且N可編碼成0100，其中N係未知的鹼基識別。在一些實施方案中，經編碼資料亦可包括表示MAPQ分數、片段號碼、序列長度、SAM旗標、片段的鹼基識別或核苷酸、MAPQ分數以外之用於片段的一或多個品質指示器、或其任何組合的資料。經編碼片段資料的範圍可

從描述片段的16位元的值至64位元的值，或更多。輸入單元131可將經產生片段寫入至DRAM 160。

【0057】 控制機140可偵測原始輸入資料的接收、啟動輸入單元131、及初始化對應於待基於原始輸入資料產生之K-mer圖之實例的圖描述資料。啟動輸入單元131可包括控制機140將指示輸入單元131在提供為至輸入單元131之輸入的原始資料上執行由輸入單元131之硬體邏輯電路系統定義之操作的一或多個控制訊息發送至輸入單元131。在一些實施方案中，啟動硬體邏輯單元（諸如輸入單元131）亦可包括控制機將可用以針對執行硬體邏輯單元之操作組態其的圖描述資料提供至硬體邏輯單元。組態硬體邏輯單元可包括，例如，提供指向儲存K-mer、K-mer節點的快取記憶體儲存位置的指標器；提供描述K-mer長度的資訊、或硬體邏輯單元需要於其上操作的類似者。

【0058】 初始化圖描述資料可包括，例如，控制機140產生原始輸入資料的K-mer圖識別符、圖狀態資訊資料結構的產生、或其組合。K-mer圖識別符包括一或多個字元、一或多個數字、或其組合的資料字串，該資料字串可用以在從原始圖資料由輸入單元131接收的時間至至少在特定的原始圖資料組的K-mer圖的產生完成之後使用抹除單元138將相關於K-mer圖識別符的資料從快取記憶體150、DRAM 160、或二者移除的時間的K-mer圖產生程序各處識別K-mer圖的實例。在一些實施方案中，K-mer圖識別符可包括數字，諸如具有在0至63之間的值的6位元的數字。在一些實施方案中，K-mer圖識別符甚至可用以在抹除單元用以從快取記憶體150、DRAM 160、或二者移除前述資料之後參考至K-mer圖。圖狀態資訊資料結構係具有儲存資料的一或多個欄位的資料結構，該資料描述待針對特定的原始輸入資料組產生之K-mer圖之實例的目前狀態。狀

態資訊可包括，例如，指示在用於K-mer圖之特定實例的原始圖資料上操作的最末硬體邏輯單元的資料、指示最末硬體邏輯單元是否中止操作的資料、指示K-mer長度的資料、指示K-mer節點列表的資料、指示可用以識別快取記憶體中之K-mer節點的指標器之列表的資料、K-mer節點列表的長度、指示非獨特K-mer之列表的資料、指示原始輸入資料在快取記憶體或DRAM中之位置的資料、指示K-mer圖之實例的節點在DRAM中的基底位址的資料、或其任何子集或組合。

【0059】 輸入單元131可格式化輸入片段112及參考基因體122，並將輸入片段112及參考基因體寫入至DRAM 160。控制機140可偵測輸入單元131何時已完成輸入片段112及參考基因體122的格式化及至DRAM 160的寫入。在藉由控制機140偵測到輸入片段112及參考基因體122的格式化及至DRAM的寫入完成時，控制機140可更新圖狀態資訊以指示輸入單元131已在用於K-mer圖之第一實例的第一原始圖資料上完成其操作。

【0060】 一旦第一原始圖資料已輸入、格式化、並儲存在DRAM 160中，控制機140可判定待接著啟動及組態的次一硬體邏輯單元。例如，控制機140可啟動及組態圖節點單元132，以基於經格式化及儲存在DRAM 160中的參考基因體122及片段之堆積的部分產生K-mer節點。

圖節點單元

【0061】 控制機140可藉由處理儲存在DRAM中的經格式化片段112及參考基因體122而啟動及組態圖節點單元132以繼續K-mer圖之第一實例的產生。此可包括，例如，將控制信號發送至圖節點單元132、提供圖描述資料至圖節

點單元132、或其組合。圖描述資料可用以針對操作組態圖節點單元132。例如，提供圖描述資料至圖節點單元132且控制機140可自其組態圖節點單元132以識別由圖描述資料定義之特定尺寸的K-mer。本文描述之圖描述資料的其他欄位可用以以類似方式組態硬體邏輯單元，諸如圖節點單元132。

【0062】 額外地，以實質平行的方式，控制機140可偵測輸入單元131接收到作為輸入的第二原始圖資料。控制機140接著可啟動輸入單元131、指示單元131格式化第二原始圖資料的片段及參考基因體、及產生用於待基於第二原始圖資料產生之K-mer圖之第二實例的第二圖描述資料。因此，控制機140可藉由同時管理在不同的處理階段正在不同的原始圖資料組上執行K-mer圖產生程序的不同硬體邏輯單元131、132的執行而達成高生產量位準。控制機140經組態以橫跨硬體邏輯單元131、132、133、134、135、136、137、138之各者管理此平行功能性，如此在時間上的任何特定點可有多達八個硬體邏輯單元在八個不同的原始圖資料組上操作，其中硬體加速圖產生單元130努力達到同時產生八個不同的K-mer圖。控制機140使用圖描述資料藉由啟動及組態各各別硬體邏輯單元以抽象地達成由在各各別硬體邏輯單元之間不具有直接及實體輸入/輸出連接的非管線化硬體邏輯單元131、132、133、134、135、136、137、138構成的高階管線化功能性而管理此程序生產量。雖然說明在相同時間執行八個同時K-mer圖產生的實例，本揭露可經組態以達成更多的同時K-mer圖產生，諸如藉由一次實施多個硬體加速圖產生單元130、在一或多個硬體加速圖產生單元130上的多個硬體邏輯單元的多個實例、或其組合。

【0063】 圖節點單元132可分析片段112之堆積的各片段，以識別片段之K-mer之各者。此可包括，例如，沿著各片段的各位置滑動K-mer存取窗以識別

各別片段的各特定K-mer。圖節點單元132可將表示用於各片段之各經識別K-mer的K-mer圖之節點的資料儲存在快取記憶體150中。同樣地，圖節點單元132亦可產生節點指標器之資料結構中的節點指標器的列表並儲存在DRAM中，其中各節點指標器指向K-mer節點快取記憶體位置。圖節點單元132亦可產生及儲存指示用於由控制機維持之圖描述資料中之各K-mer圖的節點指標器之列表的位置及長度的資訊。此等指標器可由控制機140使用為圖狀態資訊以在K-mer圖產生程序的後續部分期間組態另一硬體邏輯單元。快取記憶體150可採用一或多個快取一致性政策，諸如經組態以驅逐快取記憶體150的最舊物件的LRU快取一致性政策。其中最舊物件係基於該物件寫至快取記憶體150的時間判定。

【0064】 參照圖4顯示由圖節點單元132產生且儲存在快取記憶體150、DRAM 160、或二者中之資料的實例。圖4顯示參考基因體410、片段420的一部分，並提供德布魯因圖400。於下文更詳細地描述的德布魯因圖400包括用於在基因體410及片段420的該部分中的各K-mer的節點及在各對K-mer節點之間鏈接具有k-1個重疊核苷酸的一對節點的邊緣。

【0065】 參照圖4的實例，圖節點單元132可基於參考基因體410及片段420的一部分的接收產生表示德布魯因圖400之第一路徑430的節點431、432、433、434、435、436、437、438的資料，並可產生節點441、442、443、444。首先，如圖4所示，圖節點單元132可對準參考基因體410a、410b的重疊部分及片段420a、420b的重疊部分以識別重疊區域。圖節點單元132可識別基因體410及片段420之該部分的K-mer的各者。此可藉由在參考基因體410之該部分的第一位置使用長度k的存取窗擷取由存取窗識別的K-mer，該長度在此實例中等於4、產生表示包括經擷取K-mer之圖的節點的資料、將表示節點的資料儲存在快

取記憶體150中、使存取窗前進一個核苷酸、並接著迭代地重複此程序而達成。在此實例中，圖節點單元132可識別參考基因體410的該部分的K-mer ATCG、TCGC、CGCC、GCCT、CCTA、CTAG、TAGA、及AGAA，並產生各別節點431、432、433、434、435、436、437、438，其中此等節點的一者對應於各別K-mer。將各節點產生成具有k的長度，在此實例中係4，且與次一相鄰節點具有k-1個K-mer的重疊數目。圖節點單元132可將經產生節點儲存在快取記憶體150、DRAM 160、或二者中。在一些實施方案中，快取記憶體可包括散列表快取記憶體。在此類實施方案中，可將節點儲存為散列表的鍵。圖節點單元132可儲存描述指向由控制機140維持之圖描述資料中之K-mer節點位置的指標器的資料。

【0066】 圖節點單元132可對片段420執行相同操作。關於片段420，圖節點單元132可識別K-mer ATCG、TCGC、CGCG、GCGT、CGTA、GTAG、TAGA、及AGAA。此可類似地藉由在片段420的該部分的第一位置使用長度k的存取窗擷取由存取窗識別的K-mer，該長度在此實例中等於4、並接著使存取窗前進且重複該程序而達成。圖節點單元132可藉由識別基因體410之部分的各K-mer而開始。在一些實施方案中，圖節點單元132可產生K-mer之各者的對應節點。在其他實施方案中，圖節點單元132可僅產生對應於與參考基因體410之該部分的K-mer節點不同之經識別K-mers的431、432、433、434、435、436、437、438。在各場景中，將各節點產生成具有k的長度，在此實例中係4，且與次一相鄰節點具有k-1個K-mer的重疊數目。此可繼續至產生用於參考基因體410之該部分之各K-mer的節點並儲存在快取記憶體150或DRAM 160中為止。

【0067】 在一些實施方案中，圖節點單元132亦可經組態以識別非獨特K-mer。在此類實施方案中，圖節點單元132可對片段之第一堆積的各特定片段判定經識別K-mer是否係獨特K-mer或非獨特K-mer。若圖節點單元132判定特定K-mer係獨特K-mer，則圖節點單元132可使存取窗前進單一個核苷酸以評估次一K-mer。替代地，若圖節點單元132判定特定K-mer係非獨特K-mer，則圖節點單元132可儲存指示該特定K-mer係非獨特K-mer的資料。例如，圖節點單元132可針對K-mer圖的特定實例將指示K-mer係非獨特K-mer的資料旗標儲存在由控制機維持的圖描述資料中。然而，此類資料可由硬體加速圖產生單元130的任何其他組件儲存、儲存在硬體加速圖產生單元130的任何其他記憶體單元、或其組合中。接著，後續的硬體邏輯單元可執行解決非獨特K-mer的操作，以降低或消除K-mer圖之實例中的循環。

【0068】 在程序中的此點，圖節點單元132將表示各K-mer之K-mer圖的節點的資料儲存在快取記憶體150、DRAM 160、或二者中。亦即，硬體加速圖產生單元130尚未產生圖邊緣431a、432a、433a、434a、435a、436a、437a、432b、441a、442a、443a、444a、圖邊緣權重、或類似者。K-mer圖之此實例的此等特徵可藉由硬體加速圖產生單元130的一或多個其他硬體邏輯單元產生。

【0069】 控制機140可監測圖節點單元132的操作。一旦圖節點單元132產生表示用於K-mer圖之此第一實例的第一原始圖資料的各片段的各K-mer的K-mer節點的資料，控制機可更新圖描述資料以指示硬體加速圖產生單元130已在第一原始圖資料上完成圖節點單元132的操作。額外地，控制機140亦可儲存包括，例如，識別非獨特K-mer之各者的旗標、K-mer圖節點的儲存位置、及指示圖節點單元132已完成其操作之資料的圖描述資料。

【0070】 一旦K-mer圖節點已產生並儲存在快取記憶體150中，控制機140可判定待接著啟動及組態的次一硬體邏輯單元。例如，控制機140可啟動並組態圖邊緣單元133以產生權重、或二者、節點對之間的圖邊緣。

圖邊緣單元

【0071】 圖邊緣單元133可產生在由圖節點單元132產生之K-mer節點對之間的圖邊緣。一旦判定用於第一K-mer圖實例的圖節點單元132已完成且圖邊緣單元133可用，控制機140可啟動圖邊緣單元133。在一些實施方案中，控制機140可將儲存由圖節點單元132為K-mer圖的特定實例產生的K-mer圖節點的位置提供給圖邊緣單元133或以其他方式使該圖邊緣單元可存取該等位置。例如，控制機140可在用於K-mer圖之實例的K-mer節點的產生期間存取由圖節點單元132產生及儲存的K-mer圖描述資料。經存取的K-mer圖描述資料可指示或以其他方式描述K-mers的列表。

【0072】 一旦圖邊緣單元133已獲得用於K-mer圖之第一實例的K-mer圖節點的位置，圖邊緣單元可開始產生在表示K-mer節點的資料之間的一或多個圖邊緣。在一些實施方案中，圖邊緣單元133可從散列表快取記憶體存取表示用於特定片段之K-mer之各者的圖節點的資料。圖邊緣單元133可產生用於儲存在散列表快取記憶體中之表示在用於K-mer的圖節點之間的圖邊緣的資料。例如，表示圖邊緣的資料可作為用於邊緣之來源節點的圖節點記錄的部分儲存在散列表快取記憶體中。在一些實施方案中，圖邊緣單元133可將邊緣權重指派給K-mer圖的各邊緣。例如，圖邊緣單元133可對鏈接各別K-mer對之圖邊緣的每個發生增加+1，或其他權重。

【0073】 舉實例而言，圖邊緣單元133可使用從控制機140獲得的圖控制資料識別係相鄰節點的節點。可基於各種因素將節點判定成係相鄰節點，包括節點共用 $k-1$ 個重疊核苷酸及節點在滑動 K -mer存取窗的二個連續位置中觀察到的判定。例如，圖邊緣單元133可沿著原始圖資料之各片段的各位置滑動 K -mer存取窗。在一些實施方案中，圖邊緣單元133可在判定在片段中觀察到二個連續 K -mer（除了一個鹼基外，其等完全重疊）時產生邊緣或遞增邊緣權重。圖節點133在此一場景中產生邊緣或遞增邊緣權重，因為此場景暗示邊緣在對應於該等二個連續 K -mer的圖節點之間。

【0074】 在一些實施方案中，可將表示圖節點的資料儲存為散列表的散列鍵。在此類實施方案中，圖邊緣單元133可藉由存取第一節點（或散列鍵）所映射的散列位置並產生用於儲存在該散列位置之指向第二節點（或散列鍵）的指標器而產生從第一節點（或散列鍵）至第二節點（或散列鍵）的邊緣。後續邊緣可以相同方式產生，其產生通過可使用一或多個圖游走演算法游走之圖的路徑430或440。

【0075】 參照圖4的實例，圖邊緣單元133可產生表示第一路徑430的節點431、432、433、434、435、436、437、438對、第二路徑440的節點441、442、443、444對、或第一路徑430及第二路徑440中的一或多個節點對之間的一或多個邊緣的資料。此等邊緣的實例在圖4中顯示成邊緣431a、432a、433a、434a、435a、436a、437a、432b、441a、442a、443a、444a。

【0076】 在此實例中，將核苷酸的序列420稱為片段。然而，在一些實施方案中，低品質鹼基移除可能發生，使得核苷酸的序列420係片段重疊組或片段的部分。在此類實施方案中，鏈接一對節點的圖邊緣將僅建立在特定片段重

疊組內且不從第一片段重疊組的K-mer節點至第二組態的K-mer節點的一或多個鏈接的路徑。片段重疊組可包括在移除低品質鹼基後發生的核苷酸序列。

反向傳播單元

【0077】 硬體加速圖產生單元130可包括反向傳播單元134。然而，控制機140在某些實施方案中可僅啟動及組態反向傳播單元134。例如，當由控制機140判定由目前的原始圖資料組產生的K-mer圖稍後將轉變成序列圖時，控制機140可啟動反向傳播單元134。當啟動時，反向傳播模組134可接收來自控制機140的圖描述資料。在此類實施方案中，當由序列圖繼承時，對圖邊緣權重的調整可使權重更可靠。

【0078】 圖邊緣單元133可藉由識別快取記憶體150中的對應K-mer節點、建立鏈接K-mer節點的邊緣、且接著對各個發生將邊緣的權重遞增+1而在片段重疊組的K-mer節點之間建立邊緣。在一些實施方案中，在片段重疊組的K-mer節點已使用圖邊緣單元134加至K-mer圖並加權後，反向傳播單元134可用以通過K-mer圖中的片段重疊組的開始節點的圖邊緣「左側」的線性鏈將+1邊緣權重增量反向傳播k-1個階段。在此上下文中，K-mer圖中之片段重疊組的開始節點的「左側(left)」係與K-mer圖之引導邊緣相對之K-mer圖的方向。為說明此概念，德布魯因圖400中之節點434的「左側」將係節點433、432、及431。

【0079】 例如，在一些實施方案中，反向傳播模組134可在散列表快取記憶體中存取表示包括該K-mer圖的K-mer節點、對應的圖邊緣、或類似者之K-mer圖的資料，且接著對在新組態開始之前發生的K-1個節點調整邊緣權重。反向傳播單元135可定位適當的K-mer節點及邊緣，以基於從控制機160接收之包

括指向儲存此資訊之快取記憶體位置之指標器的圖描述資料調整。圖描述資料可使用在反向傳播期間發生的任何變化更新。

【0080】 在一些實施方案中，執行上文提及的反向傳播可能係有利的，因為N-鹼基片段重疊組可僅遞增一系列的(N-K)個邊緣權重。然而，若稍後將此K-mer圖轉變成對應於此片段重疊組之具有(N-1)個內部邊緣的序列圖，則第一(K-1)個邊緣權重不會適當地繼承經遞增邊緣權重。上文提及的反向傳播處理此問題的大部分實例。因此，反向傳播可用以處理此問題，以在將K-mer圖變換成序列圖時增加所繼承之邊緣權重的可靠性。

循環單元

【0081】 硬體加速圖產生單元130可包括循環單元135。在一些實施方案中，循環單元135可由控制機140啟動及組態以偵測K-mer圖之實例中的循環。例如，循環單元135可評估已由輸入單元131、圖節點單元132、圖邊緣單元133、及反向傳播單元134的一或多者產生之K-mer圖之實例的原始圖資料的K-mer節點及K-mer邊緣。循環單元135經組態以接收來自控制機140的圖描述資料。循環單元135可迭代地標記用於刪除的頭節點。頭節點可包括不包括任何內接邊緣(in-edges)的節點。其中內接邊緣係從第一節點指向至該第一節點自身的邊緣。在標記用於刪除的各頭節點後，循環單元135可判定由外接邊緣所指向的任何節點是否已由於經刪除節點而變成頭節點。若判定此類節點，標記其等以用於刪除。外接邊緣係從第一節點指向另一節點的圖邊緣。循環單元135可繼續執行此程序，直到沒有頭節點留下為止。

【0082】 在判定沒有頭節點留下後，循環單元135可判定圖是否係空的，其中空圖意指該圖的所有節點均針對刪除標記。若無頭圖係空的，則沒有循環。替代地，若無頭圖係非空的，則該圖必須含有循環。因為沒有循環中的節點藉由刪除循環外側的節點而變成頭節點，循環抗拒此類型的刪除。

【0083】 在作出此等判定中的其中一者後，循環單元135可對控制機140提供是否偵測到循環的指示。接著，基於從循環單元135提供的指示，控制機140可判定接著應啟動及組態哪個硬體邏輯單元。例如，若偵測到循環且應中止K-mer圖之實例的產生，則控制機140可啟動並組態抹除單元138。在此類情況中，抹除單元可從快取記憶體150及DRAM 160刪除對應於經中止K-mer圖之實例的原始圖資料。替代地，若K-mer圖之實例的產生繼續，則控制機140可啟動並組態另一硬體邏輯單元以在原始圖資料上執行後續操作以產生K-mer圖的實例。例如，若K-mer圖產生繼續，控制機140可啟動並組態修剪單元136或圖輸出單元137的其中一者。

【0084】 雖然循環單元135可由硬體加速圖產生單元130使用以偵測正在產生之K-mer圖之實例中的循環，循環單元135可如反向傳播單元134般地選擇性地啟動及組態。此係因為可預見一些類型的K-mer圖可包括循環。然而，對於某些種類的K-mer圖，不具有具有循環的圖可能係有益的。因此，硬體加速圖產生單元130可針對，例如，控制機140組態以接收指示循環單元135是否應對圖的特定實例執行的輸入。

修剪單元

【0085】 硬體加速圖產生單元130可包括修剪單元135。如同反向傳播單元134及循環單元135，修剪單元136可選擇性地由控制機160啟動及組態。若經啟動及組態，修剪單元135可評估在用於已於此點由硬體加速圖產生單元130產生之K-mer圖之實例的原始圖資料中的各圖邊緣的權重。在一些實施方案中，若修剪單元136判定圖邊緣的權重值無法滿足預定臨限，則修剪單元136可刪除圖邊緣及在經識別圖邊緣之後發生的任何K-mer節點。替代地，若修剪單元136判定圖邊緣的權重值滿足預定臨限，則修剪單元136將保持圖邊緣完整。

【0086】 在其他實施方案中，修剪單元136可識別線性鏈，其中線性鏈係通過圖的最大路徑，其中開始節點與結束節點之間的所有內部節點恰具有一個內接邊緣及一個外接邊緣。在此類實施方案中，修剪單元136可判定線性鏈的每個內部邊緣是否均無法滿足修剪臨限。若此一場景發生，修剪單元136可刪除除了鏈的開始節點及/或結束節點以外的整體線性鏈（包括所有內部邊緣及所有內部節點），若該開始節點及/或結束節點具有任何非內部邊緣，修剪單元136可保留其。

圖輸出單元

【0087】 硬體加速圖產生單元130可包括圖輸出單元137。當K-mer圖的實例係由圖描述資料及經快取資料描述時，圖輸出單元137可由硬體加速圖產生單元130用以產生K-mer圖之實例的最終版本170。例如，圖輸出單元137可使用包括，例如，指向散列表快取記憶體中之儲存K-mer圖資料之位置的指標器的圖描述資料從散列表快取記憶體150獲得表示K-mer圖的資料。接著，圖輸出

單元137可將從散列表快取記憶體150獲得之描述K-mer圖170之最終版本的資料提供至變異識別單元180。變異識別單元180可在K-mer圖170的最終版本上執行變異識別分析以產生一組變異190。一組變異190可包括一或多個候選變異。變異係生物體之基因資料中的改變。候選變異係由變異識別單元基於處理K-mer圖170推斷之藉由變異識別單元的判定。在一些實施方案中，候選變異在變異判定時可具有臨限誤差位準。

【0088】 變異識別單元180可藉由處理K-mer圖180而識別候選變異。在一些實施方案中，例如，當鹼基識別或核苷酸在片段之堆積的一或多個片段與在參考基因體的特定位置之參考基因體的核苷酸不同時，變異識別單元180可識別候選變異。描述該組變異190的資料可以任何數目的方式產生或判定。例如，在一些實施方案中，變異識別操作可如在，例如，美國專利公開案第2016/0180019號、美國專利公開案第2016/0306922號、及美國專利公開案第2019-0259468號中所更詳細地描述地執行，其等各者之全部內容特此以引用方式併入本文中。描述該組變異190的資料可以若干不同的方式提供以用於輸出。例如，描述該組變異190的資料可顯示在核酸定序儀110的顯示器上、顯示在不同電腦的顯示器上、經由計算裝置的一或多個揚聲器音訊地輸出、經由印表機、或任何其組合輸出。

抹除單元

【0089】 抹除單元138可用以在藉由硬體加速圖產生單元130完成及輸出K-mer圖之實例時執行記憶體回收任務。例如，抹除單元可刪除與藉由硬體加速圖產生單元130完成及輸出之K-mer圖之特定實例有關的所有原始圖資料。替

代地或額外地，抹除單元138可刪除與藉由控制機儲存之圖形單元130的特定實例相關的所有資料、刪除與儲存在DRAM 160或類似者中之圖形單元130的特定實例相關的所有資料。因此，抹除單元138可選擇性地從散列表快取記憶體刪除表示K-mer圖之圖節點的資料及表示K-mer圖之圖邊緣的資料。此類刪除係選擇性的，因為只有快取記憶體、控制機、或DRAM的一部分內容需要移除。另外，當將圖儲存為散列表時，該散列表可能常係稀疏地填充的，且對於抹除單元138，僅選擇性地抹除經佔據散列表項目比清除整體散列表更快，因此導致效能改善。

【0090】 然而，在一些實施方案中，與圖關聯的非散列表資料不需要逐項抹除。在此類實施方案中，抹除單元138可將圖描述資料中的列表長度設定成零，或可在不抹除內容的情況下僅將經配置記憶體空間釋放以供重使用。

【0091】 圖2係用於硬體加速產生K-mer圖之程序200的實例的流程圖。通常而言，程序200可包括獲得第一組核酸序列，其中該第一組核酸序列包括(i)對應於參考序列之作用區域的複數個片段，及(ii)參考序列的一部分(210)、使用所獲得之第一組核酸序列並使用可程式化邏輯裝置的複數個非管線化硬體邏輯單元產生K-mer圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的不同硬體邏輯電路，其中K-mer圖的各節點表示K-mer，該圖的各邊緣表示一對K-mer之間的鏈接，且K-mer圖之各邊緣的各權重表示藉由一對K-mer表示的K-mer序列的發生數目(220)、及在該K-mer圖的產生期間：在由用以產生K-mer圖的至少一部分的各硬體邏輯單元執行一或多個操作之後，使用控制機週期地更新用於K-mer圖的圖描述資料，其中圖描述資料表示(i)K-mer圖識別符，及(ii)K-mer圖狀態資訊，其中該控制機藉由在K-mer圖的產生期間觸發各各別硬

體邏輯單元的一或多個操作的執行而使用非管線化硬體邏輯單元建立操作的工作流程(230)、及提供K-mer圖至變異識別模組，其中變異識別單元處理K-mer圖以判定複數個片段的一或多者與參考序列之間的候選變異。

【0092】 圖3係用於硬體加速產生K-mer圖之程序300的另一實例的流程圖。通常而言，程序300可包括獲得第一組核酸序列，其中第一組核酸序列包括(i)對應於參考序列之作用區域的複數個片段，及(ii)參考序列的一部分(310)、對於第一組核酸序列的各特定核酸序列：針對儲存在散列表快取記憶體中並藉由第一硬體邏輯單元產生表示用於特定核酸序列之各K-mer的圖節點的資料(320)、藉由控制機偵測第一硬體邏輯單元已完成用於特定核酸序列之各K-mer的圖節點的產生(330)、藉由控制機組態第二硬體邏輯單元以執行所產生之圖節點的圖邊緣產生(340)、及對於一或多對所產生之圖節點：藉由第二硬體邏輯單元並用於儲存在圖散列表中，產生表示藉由第一硬體邏輯單元產生的一或多對所產生之圖節點之間的圖邊緣的資料，其中儲存在散列表快取記憶體中之表示用於各K-mer的圖節點的資料及儲存在散列表快取記憶體中之表示圖邊緣的資料表示第一組核酸序列的K-mer圖(350)。

【0093】 圖4係K-mer圖400的實例。在此實例中，K-mer圖係400基於參考基因體410及片段420的至少一部分產生。在此實例中，K-mer圖400係德布魯因圖。

【0094】 K-mer圖400係使用複數個節點及節點對之間的一或多個邊緣產生。各節點表示長度k的K-mer，其中在此實例中k=4。各邊緣提供藉由邊緣鏈接的K-mer有k-1個核苷酸的重疊的指示。在K-mer圖400中，路徑430包括表示

參考序列410之部分的各K-mer的複數個節點及邊緣。接著，路徑440包括表示與參考基因體410的該部分不同之片段420的部分的複數個節點及邊緣。

【0095】 圖5係可用於硬體加速K-mer圖之系統500組件的實例的方塊圖。

【0096】 計算裝置500企圖表示各種形式的數位電腦，諸如，膝上型電腦、桌上型電腦、工作站、個人數位助理、伺服器、刀鋒伺服器、大型電腦、及其他適當的電腦。計算裝置550企圖表示各種形式的行動裝置，諸如個人數位助理、蜂巢式電話、智慧型手機、及其他類似的計算裝置。額外地，計算裝置500或550可包括通用串列匯流排(USB)快閃驅動器。USB快閃驅動器可儲存作業系統及其他應用程式。USB快閃驅動器可包括輸入/輸出組件，諸如可插入至另一計算裝置之USB埠中的無線傳輸器或USB連接器。此處所示的組件、其等的連接及關係、及其等功能僅意圖作為實例，且未意圖限制本文件中描述及/或主張之本發明的實施方案。

【0097】 計算裝置500包括處理器502、記憶體504、儲存裝置506、連接至記憶體504及高速擴充埠510的高速介面508、及連接至低速匯流排514及儲存裝置506的低速介面512。組件502、504、506、508、510、及512之各者使用各種匯流排互連，並可安裝於共同主機板上或依需要以其他方式安裝。處理器502可在計算裝置500內處理用於執行的指令，包括儲存在記憶體504中或儲存裝置506上的指令，以將用於GUI的圖形資訊顯示在外部輸入/輸出裝置上，諸如耦接至高速介面508的顯示器516。在其他實施方案中，多個處理器及/或多個匯流排可依需要連同多個記憶體及多種類型的記憶體使用。再者，可將多個計

算裝置500連接成，例如，伺服器庫、刀鋒伺服器群組、或多處理器系統，其中各裝置提供必要操作的部分。

【0098】 記憶體504將資訊儲存在計算裝置500內。在一實施方案中，記憶體504係（多個）揮發性記憶體單元。在另一實施方案中，記憶體504係（多個）非揮發性記憶體單元。記憶體504亦可係另一形式的電腦可讀媒體，諸如磁碟或光碟。

【0099】 儲存裝置506能夠為計算裝置500提供大量儲存。在一個實施方案中，儲存裝置506可係或含有電腦可讀媒體，諸如軟碟裝置、硬碟裝置、光碟裝置、或磁帶裝置、快閃記憶體或其他類似的固態記憶體裝置、或裝置的陣列，包括在儲存區網路中或採其他組態的裝置。電腦程式產品可有形地具現在資訊載體中。電腦程式產品亦可含有指令，當執行該等指令時，執行一或多種方法，諸如於上文描述者。資訊載體係電腦或機器可讀媒體，諸如記憶體504、儲存裝置506、或處理器502上的記憶體。

【0100】 高速控制器508為計算裝置500管理帶寬密集的操作，而低速控制器512管理較低的帶寬密集操作。此類功能的配置僅係實例。在一個實施方案中，高速控制器508，例如，經由圖形處理器或加速器耦接至記憶體504、顯示器516、並耦接至可接受各種擴充卡（未圖示）的高速擴充埠510。在實施方案中，低速控制器512耦接至儲存裝置506及低速擴充埠514。可包括各種通訊埠（例如，USB、藍牙、乙太網路、無線乙太網路）的低速擴充埠可耦接至一或多個輸入/輸出裝置，諸如，鍵盤、指標裝置、麥克風/揚聲器對、掃描器、或網路裝置（諸如交切換或路由器），例如，通過網路配接器。如圖所示，計算裝置500可以若干不同形式實施。例如，其可實施為標準伺服器520，或在此

類伺服器的群組中實施多次。其亦可實施為機架伺服器系統524的部分。額外地，其可以個人電腦實施，例如膝上型電腦522。替代地，來自計算裝置500的組件可與行動裝置（未圖示）（諸如裝置550）中的其他組件組合。此類裝置之各者可含有計算裝置500、550的一或多者，且整個系統可由彼此通訊的多個計算裝置500、550構成。

【0101】 如圖所示，計算裝置500可以若干不同形式實施。例如，其可實施為標準伺服器520，或在此類伺服器的群組中實施多次。其亦可實施為機架伺服器系統524的部分。額外地，其可以個人電腦實施，例如膝上型電腦522。替代地，來自計算裝置500的組件可與行動裝置（未圖示）（諸如裝置550）中的其他組件組合。此類裝置之各者可含有計算裝置500、550的一或多者，且整個系統可由彼此通訊的多個計算裝置500、550構成。

【0102】 計算裝置550包括處理器552、記憶體564、及輸入/輸出裝置，尤其係諸如顯示器554、通訊介面566、及收發器568。裝置550亦可具備儲存裝置，諸如微型硬碟或其他裝置，以提供額外儲存。組件550、552、564、554、566、及568之各者使用各種匯流排互連，且該等組件的數者可安裝於共同主機板上或依需要以其他方式安裝。

【0103】 處理器552可在計算裝置550內執行指令，包括儲存在記憶體564中的指令。處理器可實施為包括分開及多個類比及數位處理器之晶片的晶片組。額外地，處理器可使用若干架構的任一者實施。例如，處理器510可係CISC（複雜指令集電腦）處理器、RISC（精簡指令集電腦）處理器、或MISC（最小指令集電腦）處理器。處理器可提供，例如，裝置550之其他組件的協

調（諸如使用者介面、由裝置550運行之應用程式、及藉由裝置550之無線通訊的控制）。

【0104】 處理器552可通過耦接至顯示器554的控制介面558及顯示介面556與使用者通訊。顯示器554可係，例如，TFT（薄膜電晶體液晶顯示器）顯示器或OLED（有機發光二極體）顯示器、或其他適當的顯示技術。顯示介面556可包含用於驅動顯示器554以呈現圖形及其他資訊給使用者的適當電路系統。控制介面558可接收來自使用者的命令並轉換其等以用於提交至處理器552。額外地，可提供與處理器552通訊的外部介面562，以使裝置550能與其他裝置近場通訊。外部介面562可，例如，在一些實施方案中提供有線通訊、或在其他實施方案中提供無線通訊，且亦可使用多個介面。

【0105】 記憶體564將資訊儲存在計算裝置550內。可將記憶體564實施為（多個）電腦可讀媒體、（多個）揮發性記憶體單元、或（多個）非揮發性記憶體單元的一或多者。擴充記憶體574亦可通過擴充介面572提供並連接至裝置550，該擴充介面可包括，例如，SIMM（單直插記憶體模組）卡介面。此類擴充記憶體574可為裝置550提供額外的儲存空間，或亦可儲存用於裝置550的應用程式或其他資訊。具體地說，擴充記憶體574可包括指令以實行或補充上述程序，且亦可包括安全資訊。因此，例如，可將擴充記憶體574提供為裝置550的安全性模組，且可使用允許裝置550的安全使用的指令程式化。額外地，安全應用程式連同額外資訊可經由SIMM卡提供，諸如以不可入侵的方式將識別資訊置於SIMM卡上。

【0106】 記憶體可包括，例如，快閃記憶體及/或NVRAM記憶體，如下文所討論的。在一個實施方案中，將電腦程式產品有形地具現在資訊載體中。

電腦程式產品含有指令，當執行該等指令時，執行一或多種方法，諸如於上文描述者。資訊載體係可，例如，透過收發器568或外部介面562接收的電腦或機器可讀媒體，諸如記憶體564、擴充記憶體574、處理器552上的記憶體。

【0107】 裝置550可通過通訊介面566無線地通訊，該通訊介面可包括必要的數位信號處理電路系統。通訊介面566可在各種模式或協定下提供通訊，尤其係諸如GSM語音通話、SMS、EMS、MMS傳訊、CDMA、TDMA、PDC、WCDMA、CDMA2000、或GPRS。此類通訊可通過，例如，射頻收發器568發生。額外地，短程通訊可，諸如使用藍牙、Wi-Fi、或其他此類收發器（未圖示）發生。額外地，GPS（全球定位系統）接收器模組570可將額外的導航及位置相關無線資料提供給裝置550，該無線資料可依需要由在裝置550上運行的應用程式使用。

【0108】 裝置550亦可使用音訊編碼解碼器560音訊地通訊，該音訊編碼解碼器可接收來自使用者的口述資訊並將其轉換成可用的數位資訊。音訊編碼解碼器560可同樣地為使用者產生可聽聲音，諸如通過，例如，裝置550之手持裝置中的揚聲器。此類聲音可包括來自語音電話通話的聲音、可包括經錄音聲音，例如，語音訊息、音樂檔案等、且亦可包括由在裝置550上操作的應用程式產生的聲音。

【0109】 如圖所示，計算裝置550可以若干不同形式實施。例如，其可實施為蜂巢式電話580。其亦可實施為智慧型手機582、個人數位助理、或其他類似行動裝置的部分。

【0110】 此處描述之系統及方法的各種實施方案可以數位電子電路系統、積體電路系統、特殊設計的ASIC（特殊應用積體電路）、電腦硬體、韌

體、軟體、及/或此類實施方案的組合實現。此等各種實施方案可包括採用一或多個電腦程式的實施方案，該一或多個電腦程式可在包括至少一個可係特殊目的或通用的可程式化處理器的可程式化系統上執行及/或解譯，該可程式化處理器耦接至儲存系統、至少一個輸入裝置、及至少一個輸出裝置以接收來自其等的資料及指令及傳輸資料及指令至其等。

【0111】 此等電腦程式（亦稱為程式、軟體、軟體應用程式、或碼）包括用於可程式化處理器的機器指令，並可以高階程序化及/或物件導向程式語言實施、及/或以組合/機器語言實施。如本文所使用的，用語「機器可讀媒體 (machine-readable medium)」、「電腦可讀媒體 (computer-readable medium)」係指用以提供機器指令及/或資料至可程式化處理器的任何電腦程式產品、設備、及/或裝置（例如，磁碟、光碟、記憶體、可程式化邏輯裝置(PLD)），包括將機器指令接收為機器可讀信號的機器可讀媒體。用語「機器可讀信號」係指用以提供機器指令及/或資料至可程式化處理器的任何信號。

【0112】 為提供與使用者的互動，此處描述的系統及技術可實施在具有用於顯示資訊給使用者的顯示裝置（例如，CRT（陰極射線管）或LCD（液晶顯示器）監視器）及使用者可藉由其提供輸入至電腦的鍵盤及指標裝置（例如，滑鼠或軌跡球）的電腦上。亦可使用其他類型的裝置以提供與使用者的互動：例如，提供給使用者的回饋可係任何形式的感覺回饋，例如，視覺回饋、聽覺回饋、或觸覺回饋；且可以任何形式接收來自使用者的輸入，包括聲波、語音、或觸覺輸入。

【0113】 此處描述的系統及技術可以包括後端組件（諸如，作為資料伺服器）、或包括中間軟體組件（例如，應用伺服器）、或包括前端組件（例

如，具有使用者可通過其與此處描述之系統及技術的實施方案互動的圖形使用者介面或網路瀏覽器的客戶端電腦）、或此類後端、中間軟體、或前端組件的任何組合的計算系統實施。系統的組件可藉由任何形式或媒體的數位資料通訊（例如，通訊網路）互連。通訊網路的實例包括區域網路（「LAN」）、廣域網路（「DI」）、及網際網路。

【0114】 計算系統可包括客戶端及伺服器。客戶端與伺服器彼此通常係遠端的，且一般通過通訊網路互動。客戶端與伺服器的關係憑藉在各別電腦上運行且對於彼此具有客戶端-伺服器關係的電腦程式而出現。

其他實施例

【0115】 已描述若干實施例。然而，將理解可在不脫離本發明之精神及範疇的情況下進行各種修改。額外地，圖式中描繪的邏輯流程不需要所示的特定順序或連續順序以達成所欲結果。額外地，其他步驟可提供至所述流程或步驟可自其消除，且可將其他組件加至所述系統或自其移除。因此，其他實施例係在下文之申請專利範圍的範疇內。

【符號說明】

【0116】

100:系統

105:生物樣本/DNA樣本

110:核酸定序儀

112:片段

- 120:參考序列資料庫/參考基因體資料庫
- 122:參考基因體/參考基因體資料庫
- 130:硬體加速圖產生單元
- 131:輸入單元/硬體邏輯單元
- 132:圖節點單元/圖單元/硬體邏輯單元
- 133:圖邊緣單元/硬體邏輯單元
- 134:反向傳播單元/硬體邏輯單元
- 135:循環單元/硬體邏輯單元
- 136:修剪單元/硬體邏輯單元
- 137:圖輸出單元/硬體邏輯單元
- 138:抹除單元/硬體邏輯單元
- 140:控制機
- 150:快取記憶體
- 160:DRAM
- 170:K-mer圖
- 180:變異識別單元
- 190:變異
- 200:程序
- 210:步驟
- 220:步驟
- 230:步驟
- 300:程序

310:步驟

320:步驟

330:步驟

340:步驟

350:步驟

400:K-mer圖

410:參考基因體

410a:參考基因體

410b:參考基因體

420:片段

420a:片段

420b:片段

430:第一路徑

431:節點

431a:邊緣

432:節點

432a:邊緣

432b:邊緣

433:節點

433a:邊緣

434:節點

434a:邊緣

435:節點

435a:邊緣

436:節點

436a:邊緣

437:節點

437a:邊緣

438:節點

440:路徑

441:節點

441a:邊緣

442:節點

442a:邊緣

443:節點

443a:邊緣

444:節點

444a:邊緣

500:系統/計算裝置

502:處理器/組件

504:記憶體/組件

506:儲存裝置/組件

508:高速介面/高速控制器/組件

510:高速擴充埠/組件

512:低速介面/低速控制器/組件
514:低速匯流排
516:顯示器
520:伺服器
522:膝上型電腦
524:機架伺服器系統
550:計算裝置/裝置/組件
552:處理器/組件
554:顯示器/組件
556:顯示介面/組件
558:控制介面
560:音訊編碼解碼器
562:外部介面
564:記憶體/組件
566:通訊介面/組件
568:收發器/組件
570:GPS（全球定位系統）接收器模組
572:擴充介面
574:擴充記憶體
580:蜂巢式電話
582:智慧型手機

【發明申請專利範圍】

【請求項1】 一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之方法，該方法包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一 K-mer 圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該 K-mer 圖的各節點表示一 K-mer，該 K-mer 圖的各邊緣表示一對 K-mer 之間的一鏈接，且該 K-mer 圖之各邊緣的各權重表示由一對 K-mer 表示的一 K-mer 序列的一發生數目；及

在該 K-mer 圖的產生期間：

在由用以產生該 K-mer 圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該 K-mer 圖的圖描述資料，其中該圖描述資料表示(i)一 K-mer 圖識別符，及(ii)K-mer 圖狀態資訊，其中該控制機藉由在該 K-mer 圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【請求項2】 如請求項1之方法，其中該複數個硬體邏輯單元的各硬體邏輯單元的該輸出係經由一散列表快取記憶體儲存。

【請求項3】 如請求項1之方法，其中該控制機係使用該可程式化邏輯裝置的一硬體邏輯單元實施。

【請求項4】 如請求項1之方法，其中該控制機係使用一或多個CPU或GPU執行軟體指令以實現該控制機的功能性而實施。

【請求項5】 如請求項1之方法，該等操作進一步包含：

將所產生之該 K-mer 圖提供至一變異識別單元，其中該變異識別單元處理該 K-mer 圖以判定該複數個片段的一或多者與該參考序列之間的候選變異。

【請求項6】 如請求項5之方法，其中藉由一或多個CPU或GPU執行軟體指令以實現該變異識別單元的一或多個功能。

【請求項7】 如請求項1之方法，其中該圖描述資料進一步包括(iii)表示該複數個硬體邏輯單元之在該K-mer圖上或在與該K-mer圖識別符關聯之該堆積的核酸序列上執行硬體邏輯的一最末硬體邏輯單元的資料。

【請求項8】 一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之系統，該系統包含：

一硬體加速圖產生單元，其包括已配置以執行操作的硬體數位邏輯電路，該等操作包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一 K-mer 圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該 K-mer 圖的各節點表示一 K-mer，該 K-mer 圖的各邊緣表示一對 K-mer 之間的一鏈接，且該 K-mer 圖之各邊緣的各權重表示由一對 K-mer 表示的一 K-mer 序列的一發生數目；及

在該 K-mer 圖的產生期間：

在由用以產生該 K-mer 圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該 K-mer 圖的圖描述資料，其中該圖描述資料表示(i)一 K-mer 圖識別符，及(ii)K-mer 圖狀態資訊，其中該控制機藉由在該 K-mer 圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【請求項9】 如請求項8之系統，其中該複數個硬體邏輯單元的各硬體邏輯單元的該輸出係經由一散列表快取記憶體儲存。

【請求項10】 如請求項8之系統，該系統進一步包含：

一或多個電腦；及

一或多個記憶體裝置，其儲存指令，該等指令當由該一或多個電腦執行時，導致該一或多個電腦執行一變異識別單元的第二操作，該等第二操作包含

藉由該變異識別單元獲得所產生之該 K-mer 圖；及

基於該變異識別單元處理所產生之該 K-mer 圖而識別一或多個候選變異，其中一候選變異係在片段之該堆積中的一或多個片段的一鹼基識別與在一參考基因體的一特定位置的該參考基因體的一核苷酸之間的一差異。

【請求項11】 如請求項8之系統，其中該等操作進一步包括：

藉由一變異識別單元獲得所產生之該 K-mer 圖；及

基於該變異識別單元處理所產生之該 K-mer 圖而識別一或多個候選變異，其中一候選變異係在片段之該堆積中的一或多個片段的一鹼基識別與在一參考基因體的一特定位置的該參考基因體的一核苷酸之間的一差異。

【請求項12】如請求項8之系統，其中該圖描述資料進一步包括(iii)表示該複數個硬體邏輯單元之在該K-mer圖上或在與該K-mer圖識別符關聯之該堆積的核酸序列上執行硬體邏輯的一最末硬體邏輯單元的資料。

【請求項13】一種用於在一可程式化邏輯裝置中硬體加速產生一K-mer圖之方法，該方法包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

對於該第一組核酸序列的各特定核酸序列：

針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各 K-mer 的一圖節點的資料；

藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各 K-mer 的一圖節點的產生；

藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及

針對一或多對所產生之該等圖節點：

藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各 K-mer 的

該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一 K-mer 圖。

【請求項14】 如請求項13之方法，該方法進一步包含：

藉由該控制機且在可由該控制機存取的一記憶體單元中週期地儲存用於該 K-mer 圖的一實例的圖描述資料，其中該圖描述資料表示(i)一 K-mer 圖識別符，及(ii)K-mer 圖狀態資訊。

【請求項15】 如請求項13之方法，

其中該第一硬體邏輯單元進一步經組態以：

判定該特定核酸序列的該等特定 K-mer 的一或多者是否與該特定核酸序列的另一 K-mer 匹配；及

基於該特定核酸序列的該一或多個特定 K-mer 與該特定核酸序列的另一 K-mer 匹配的一判定，儲存將該一或多個特定 K-mer 標記成非獨特 K-mer 的資料；

及

其中該第二硬體邏輯進一步經組態以：

將一邊緣權重指派給該 K-mer 圖的各邊緣。

【請求項16】 一種用於使用一可程式化邏輯裝置硬體加速產生一K-mer圖之系統，該系統包含：

一硬體加速圖產生單元，其包括已配置以執行操作的硬體數位邏輯電路，該等操作包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

對於該第一組核酸序列的各特定核酸序列：

針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各 **K-mer** 的一圖節點的資料；
藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各 **K-mer** 的一圖節點的產生；

藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及

針對一或多對所產生之該等圖節點：

藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各 **K-mer** 的該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一 **K-mer** 圖。

【請求項17】 如請求項16之系統，該等操作進一步包含：

藉由該控制機且在可由該控制機存取的一記憶體單元中週期地儲存用於該 **K-mer** 圖的一實例的圖描述資料，其中該圖描述資料表示(i)一 **K-mer** 圖識別符，及(ii)**K-mer** 圖狀態資訊。

【請求項18】 如請求項16之系統，

其中該第一硬體邏輯單元進一步經組態以：

判定該特定核酸序列的該等特定 **K-mer** 的一或多者是否與該特定核酸序列的另一 **K-mer** 匹配；及

基於該特定核酸序列的該一或多個特定 **K-mer** 與該特定核酸序列的另一 **K-mer** 匹配的一判定，儲存將該一或多個特定 **K-mer** 標記成非獨特 **K-mer** 的資料；

及

其中該第二硬體邏輯進一步經組態以：

將一邊緣權重指派給該 **K-mer** 圖的各邊緣。

【請求項19】 一種包括已配置以執行操作的硬體數位邏輯電路的硬體加速圖產生單元，該等操作包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

使用一可程式化邏輯裝置的複數個非管線化硬體邏輯單元使用所獲得之該第一組核酸序列產生一 **K-mer** 圖，其中各硬體邏輯單元包含經組態以執行一或多個操作的一不同的硬體邏輯電路，其中該 **K-mer** 圖的各節點表示一 **K-mer**，該 **K-mer** 圖的各邊緣表示一對 **K-mer** 之間的一鏈接，且該 **K-mer** 圖之各邊緣的各權重表示由一對 **K-mer** 表示的一 **K-mer** 序列的一發生數目；及

在該 **K-mer** 圖的產生期間：

在由用以產生該 **K-mer** 圖的至少一部分的各硬體邏輯單元執行該一或多個操作之後，使用一控制機週期地更新用於該 **K-mer** 圖的圖描述資料，其中該圖描述資料表示(i)一 **K-mer** 圖識別符，及(ii)**K-mer** 圖狀態資訊，其中該控制機藉由在該 **K-mer** 圖的產生期間觸發各各別硬體邏輯單元的該一或多個操作的執行而使用該等非管線化硬體邏輯單元建立操作的一工作流程。

【請求項20】 一種包括已配置以執行操作的硬體數位邏輯電路的硬體加速圖產生單元，該等操作包含：

獲得一第一組核酸序列，其中該第一組核酸序列包括(i)對應於一參考序列的一作用區域的複數個片段，及(ii)該參考序列的一部分；

對於該第一組核酸序列的各特定核酸序列：

針對儲存在一散列表快取記憶體中並藉由一第一硬體邏輯單元產生表示用於該特定核酸序列之各 K-mer 的一圖節點的資料；

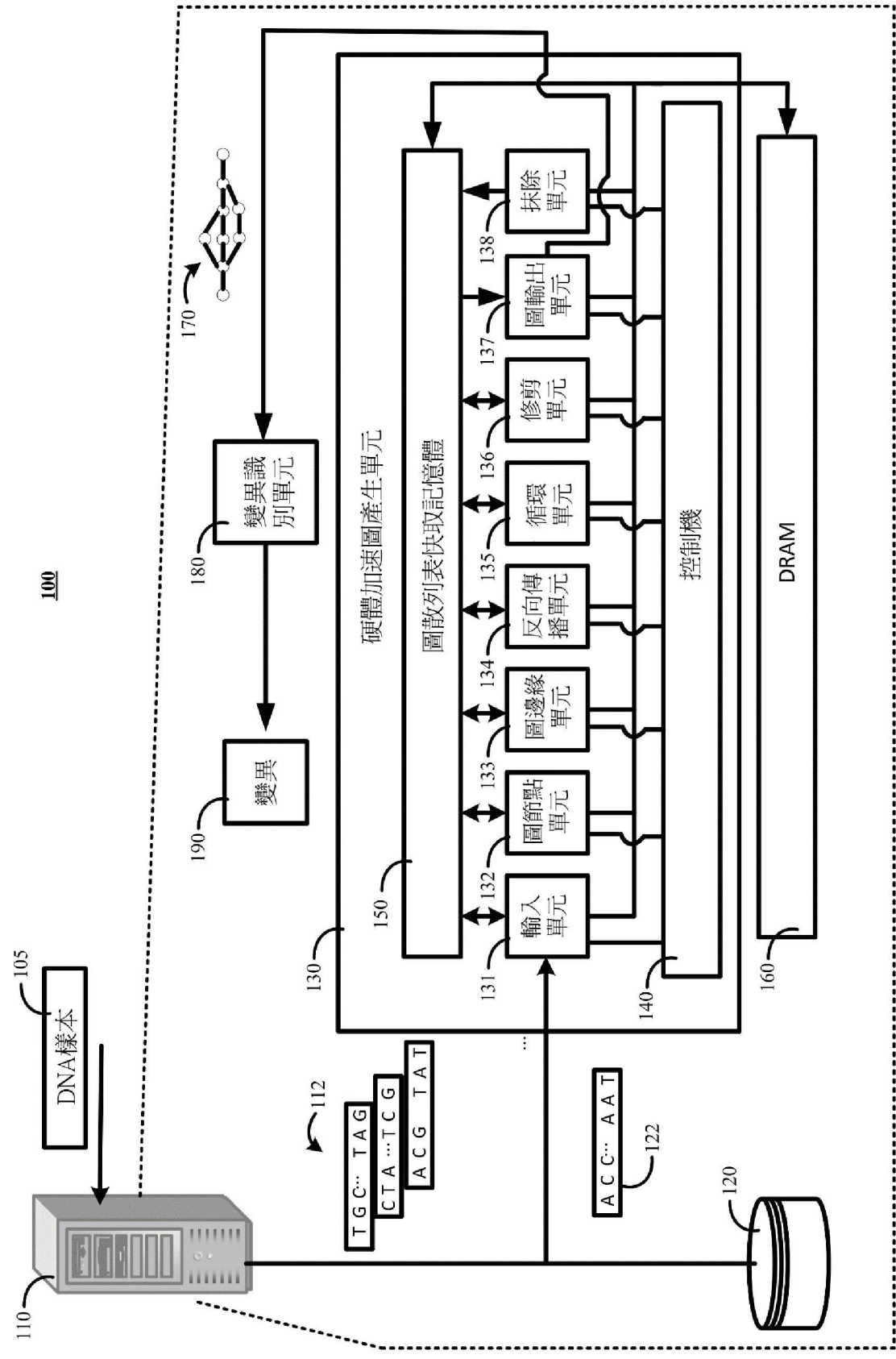
藉由一控制機偵測該第一硬體邏輯單元已完成用於該特定核酸序列之各 K-mer 的一圖節點的產生；

藉由該控制機組態一第二硬體邏輯單元以執行所產生之該等圖節點的圖邊緣產生；及

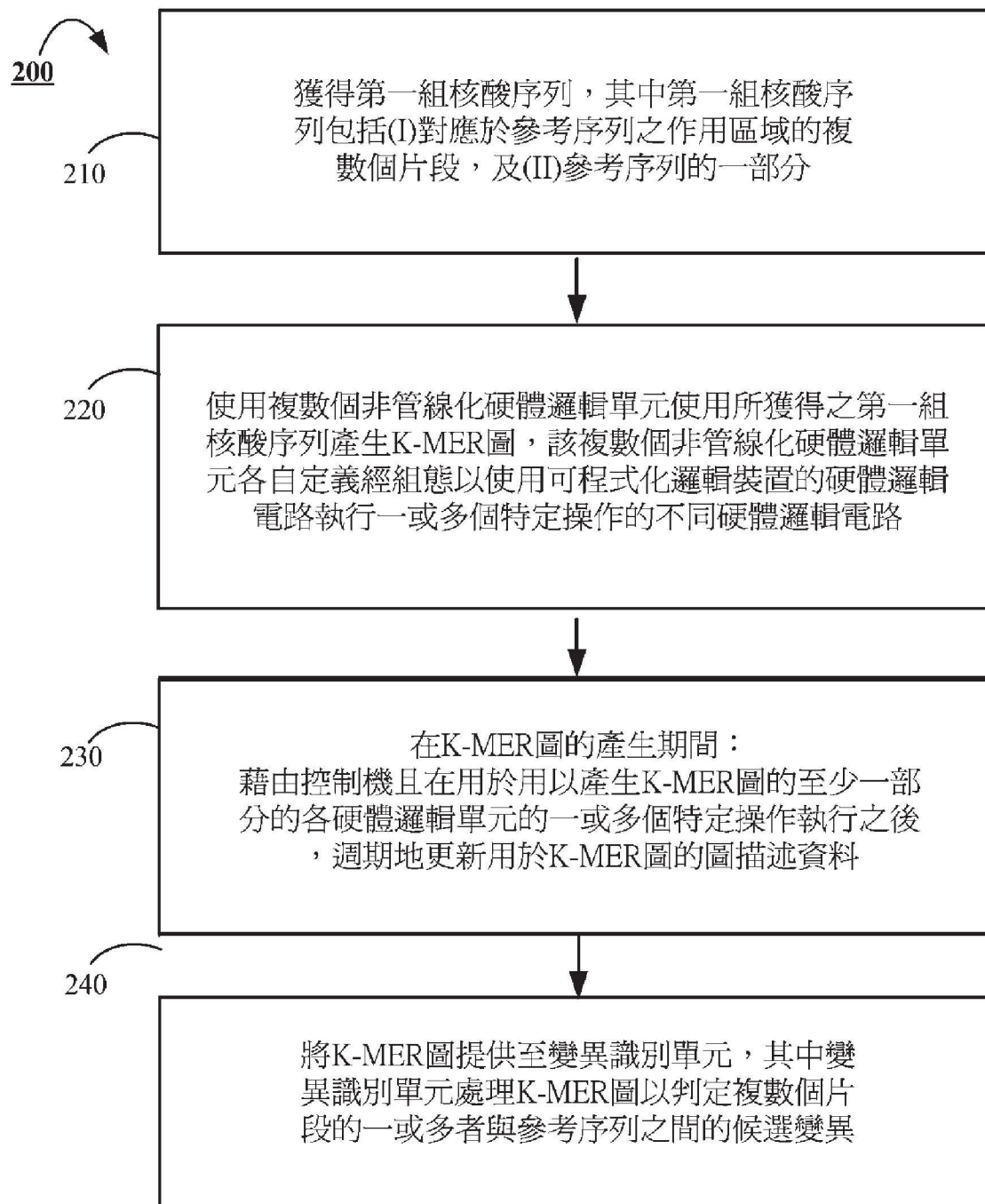
針對一或多對所產生之該等圖節點：

藉由該第二硬體邏輯單元並用於儲存在該圖散列表中，產生表示藉由該第一硬體邏輯單元產生的一或多對所產生之該等圖節點之間的圖邊緣的資料，其中儲存在該散列表快取記憶體中之表示用於各 K-mer 的該圖節點的該資料及儲存在該散列表快取記憶體中之表示圖邊緣的該資料表示該第一組核酸序列的一 K-mer 圖。

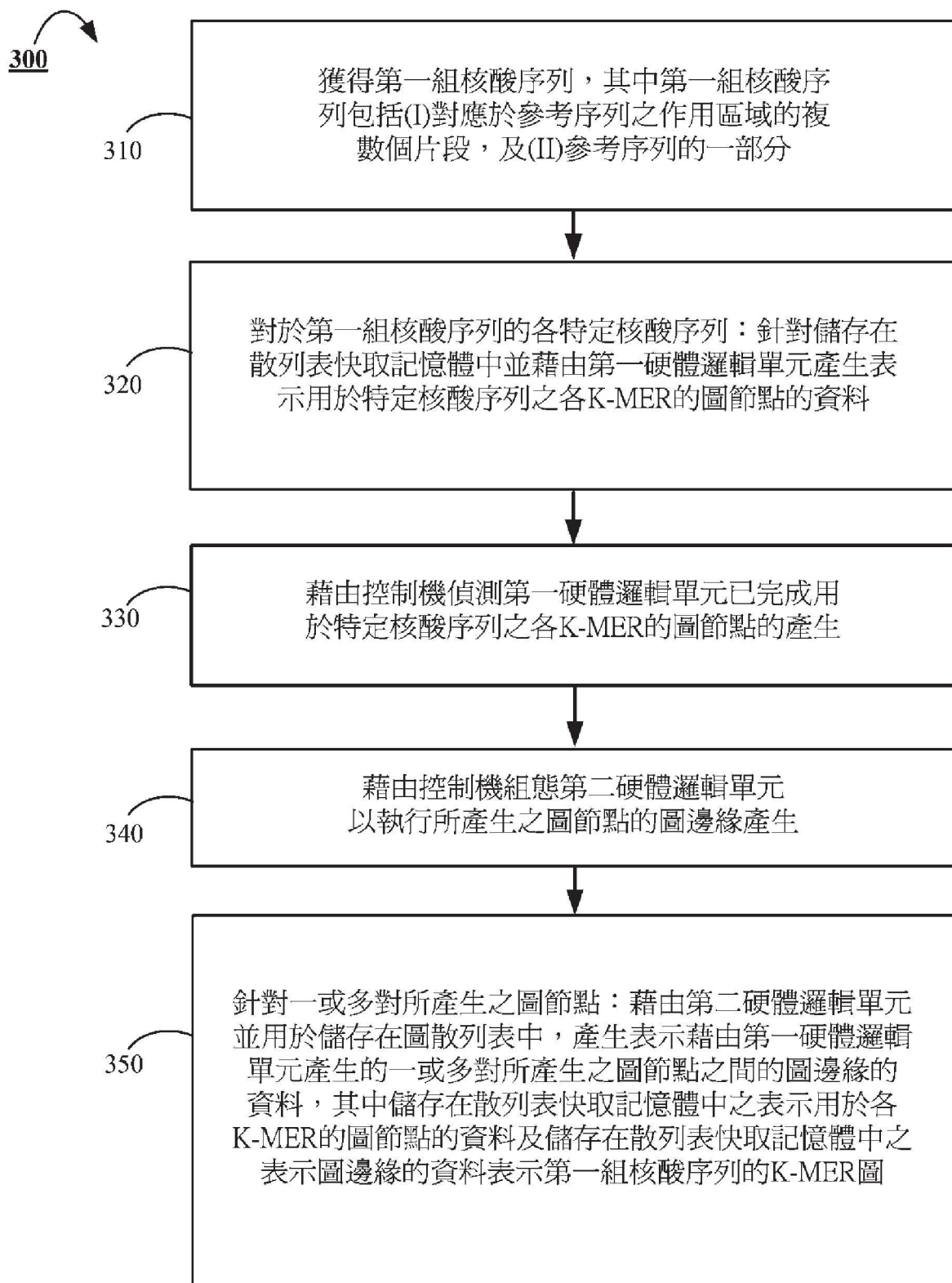
【發明圖式】



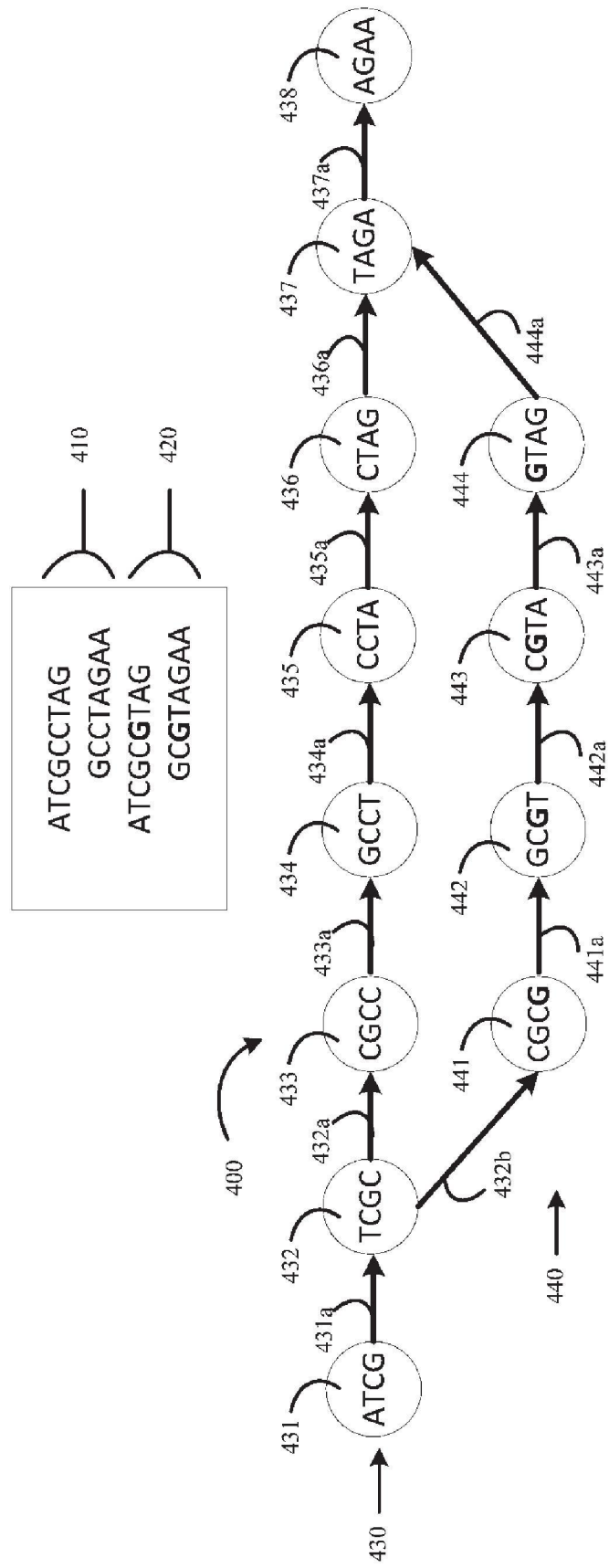
【圖 1】



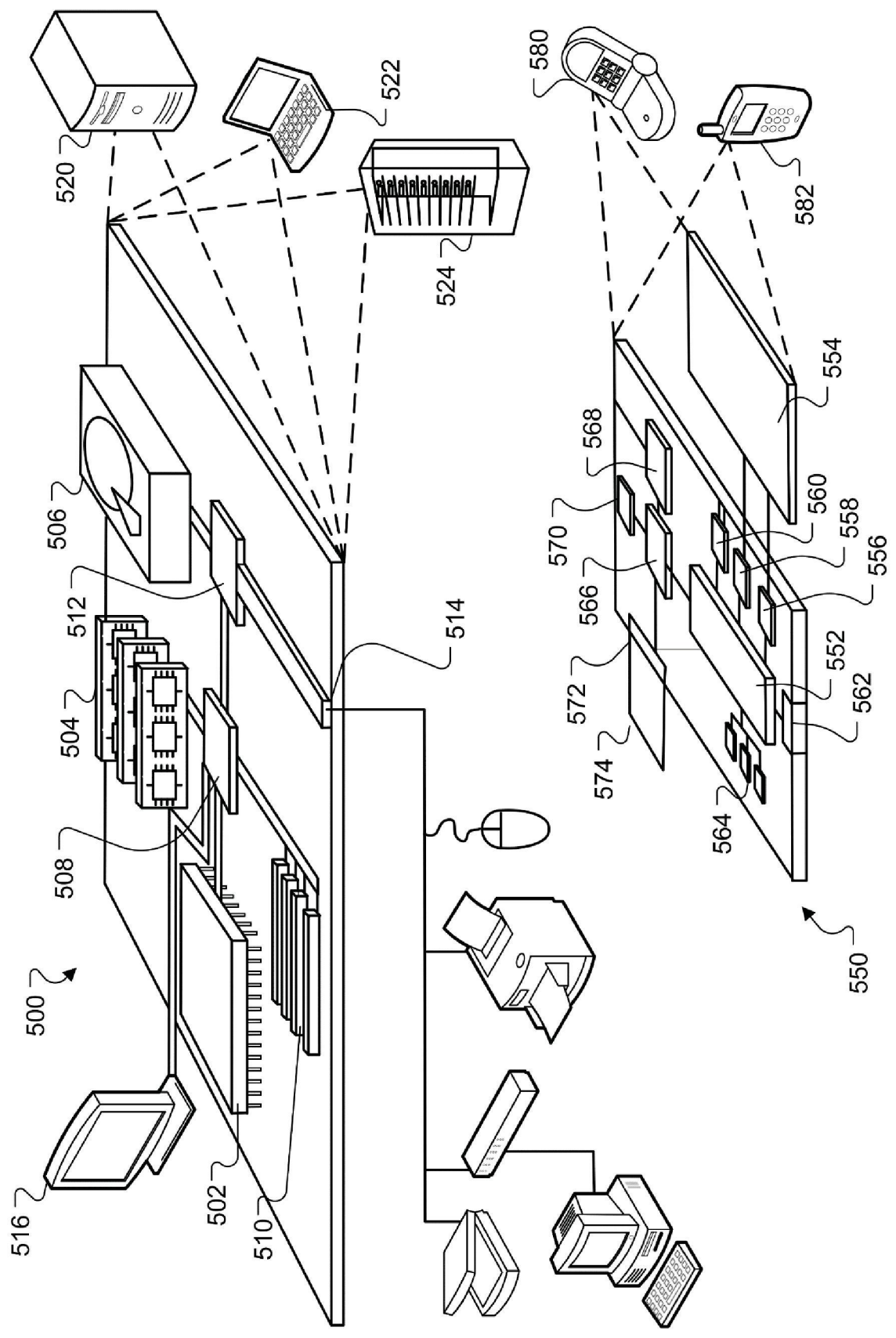
【圖 2】



【圖 3】



【圖 4】



【圖 5】