【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成31年3月28日(2019.3.28)

【公表番号】特表2018-512092(P2018-512092A)
【公表日】平成30年5月10日(2018.5.10)
【年通号数】公開・登録公報2018-017
【出願番号】特願2017-561612(P2017-561612)
【国際特許分類】
Ｇ１６Ｂ　30/00　　(2019.01)
Ｃ１２Ｑ　 1/68　　(2018.01)
【ＦＩ】
Ｇ０６Ｆ　19/22
Ｃ１２Ｑ　 1/68　　Ｚ

【手続補正書】
【提出日】平成31年2月14日(2019.2.14)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
プロセッサを含む、核酸配列情報のコンティグにスキャフォールドするためのコンピュータにより実施されるシステムであって、該プロセッサは、
初期の構成を有するコンティグ配列のセットを受け取り、
ペアエンドリードのセットを受け取り、
標準的なペアエンドリード距離度数データを受け取り、
少なくとも１つのペアエンドリード中に共存する配列を共有するコンティグペアがグループ分けされるように、コンティグペアを処理し、
コンティグを分離するためにマッピングされるリードペアのリードペア距離度数データが、初期の構成のコンティグ配列のリードペア度数データに対して標準的なペアエンドリード距離度数データにより近く近似するように、グループ化されたコンティグ配列にスキャフォールドし、ならびに、
ネットワーク、スクリーン、またはサーバーへ処理されたコンティグスキャフォールドを出力するように構成される、コンピュータにより実施されるシステム。
【請求項２】
ペアエンドリードのセットは、核酸内で内部二本鎖切断を生じさせるためにサンプルＤＮＡを消化し、複数の再連結接合部を形成するために二本鎖切断を無作為に再連結させ、および複数の再連結接合部にわたって配列決定することにより得られる、請求項１に記載のコンピュータにより実施されるシステム。
【請求項３】
標準的なペアエンドリード度数は、両方のリードが共通のコンティグにマッピングされるペアエンドリードから得られる、請求項１に記載のコンピュータにより実施されるシステム。
【請求項４】
標準的なペアエンドリード度数は、事前に生成された曲線から得られる、請求項１に記載のコンピュータにより実施されるシステム。
【請求項５】

初期の構成は無作為の構成である、請求項１に記載のコンピュータにより実施されるシステム。
【請求項６】
コンティグを分離するためにマッピングされるリードペアのリードペア距離度数データは、リードペア距離尤度が増大するときにペアエンドリード距離度数データにより近く近似する、請求項１に記載のコンピュータにより実施されるシステム。
【請求項７】
リードペア距離尤度は最大化される、請求項６に記載のコンピュータにより実施されるシステム。
【請求項８】
コンティグを分離するためにマッピングされるリードペアのリードペア距離度数データは、リードペアの距離度数データと標準的なペアエンドリード距離度数データとの間の差の統計的尺度が減少する際に、ペアエンドリード距離度数データにより近く近似する、請求項１に記載のコンピュータにより実施されるシステム。
【請求項９】
リードペア距離度数データと標準的なペアエンドリード距離度数データとの間の距離の統計的尺度は、ＡＮＯＶＡ、ｔ検定、およびＸ２乗検定の少なくとも１つを含む、請求項８に記載のコンピュータにより実施されるシステム。
【請求項１０】
コンティグを分離するためにマッピングされるリードペアのリードペア距離は、標準的なペアエンドリード距離度数と比較して得られた順序付けされたコンティグ中のリードペア距離分布の偏差が減少するときに、ペアエンドリード距離度数データとより近く一致する、請求項１に記載のコンピュータにより実施されるシステム。
【請求項１１】
標準的なペアエンドリード距離度数と比較して得られた順序付けされたコンティグ中のリードペア距離分布の偏差は最小限に抑えられる、請求項１０に記載のコンピュータにより実施されるシステム。
【請求項１２】
ペアエンドリードのセットは、核酸内で内部二本鎖切断を生じさせるためにサンプルＤＮＡを消化し、少なくとも１つの再連結接合部を形成するために二本鎖切断を再連結し、および少なくとも１つの再連結接合部にわたって配列決定することにより得られる、請求項１に記載のコンピュータにより実施されるシステム。
【請求項１３】
ＤＮＡは少なくとも１つのＤＮＡ結合剤に架橋される、請求項１２に記載のコンピュータにより実施されるシステム。
【請求項１４】
ＤＮＡは、再構成された染色質へ再度アセンブルされる分離された裸のＤＮＡである、請求項１２に記載のコンピュータにより実施されるシステム。
【請求項１５】
生体サンプルから得られる核酸配列情報のコンティグをスキャフォールドする方法であって、該方法は：
（ａ）初期の構成を持つ一連のコンティグ配列を得る工程であって、コンティグ配列は生体物質からＤＮＡを抽出することおよび生体物質を配列決定することにより得られる、工程；
（ｂ）一連のペアエンドリードを得る工程であって、一連のペアエンドリードは、核酸内に内部の二本鎖切断を生成するためにサンプルＤＮＡを消化すること、複数の再連結結合部を形成するために二本鎖切断が無作為に再連結するのを可能にすること、及び、複数の再連結結合部にわたり配列決定を行うことにより得られる、工程；
（ｃ）標準のペアエンドリードの距離度数データを得る工程；
（ｄ）少なくとも１つのペアエンドリードに共存する配列を共有するコンティグのペアを

グループ分けする工程であって、それによってグループ分けされたコンティグを生成する工程；及び、
（ｅ）コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データが、初期の構成におけるグループ分けされたコンティグのリードペア度数データに比べて少なくとも５％、標準のペアエンドリードの距離度数データとより近くに近似するように、グループ分けされたコンティグをスキャフォールドする工程、
を含む、方法。
【請求項１６】
サンプルＤＮＡが、少なくとも１つのＤＮＡ結合剤に架橋される、請求項１５に記載の方法。
【請求項１７】
サンプルＤＮＡは、再構成された染色質へ再度アセンブルされる分離された裸のＤＮＡを含む、請求項１５に記載の方法。
【請求項１８】
再構成された染色質が架橋される、請求項１７に記載の方法。
【請求項１９】
再構成された染色質がＤＮＡ結合タンパク質を含む、請求項１７に記載の方法。
【請求項２０】
再構成された染色質がナノ粒子を含む、請求項１７に記載の方法。
【請求項２１】
一連のペアエンドリードは、核酸内に内部の二本鎖切断を生成するためにサンプルＤＮＡを消化すること、複数の再連結結合部を形成するために二本鎖切断が無作為に再連結するのを可能にすること、及び、複数の再連結結合部の各側での配列決定を行うことにより得られる、請求項１５に記載の方法。
【請求項２２】
標準的なペアエンドリード距離度数データが、両方のリードが共通のコンティグにマッピングされるペアエンドリードから得られる、請求項１５に記載の方法。
【請求項２３】
標準的なペアエンドリード距離度数データが、事前に生成された曲線から得られる、請求項１５に記載の方法。
【請求項２４】
前記スキャフォールドする工程が、前記グループ分けされたコンティグの推定上の隣接するコンティグの第１のセットを選択する工程、前記リードペアに関してリードペアの距離の統計的尺度を減少させる推定上の隣接するコンティグの前記第１のセットの最小限の距離順序を判定する工程、および、リードペア距離の前記統計的尺度を減少させるために、推定上の隣接するコンティグの前記第１のセットをスキャフォールドする工程を含む、請求項１５に記載の方法。
【請求項２５】
前記最小限の距離順序を判定する工程が、推定上の隣接するコンティグの第１のセットの２つのコンティグにマッピングされているリードを含む少なくとも１つのリードペアのためのすべての可能なコンティグ構成についての予想されるリードペア距離を判定する工程を含む、請求項２４に記載の方法。
【請求項２６】
全ての可能なコンティグ構成からの他の可能な構成と比較してリードペア距離尤度を最も改善する、全ての可能なコンティグ構成からのコンティグ配向を選択する工程をさらに含む、請求項２５に記載の方法。
【請求項２７】
生体サンプルはゲノムを含む、請求項１５に記載の方法。
【請求項２８】
生体サンプルは、複数のゲノムを含むヘテロ接合サンプルである、請求項１５に記載の

方法。
【請求項２９】
Ｎ５０値は、初期の構成におけるグループ分けされたコンティグに比べて少なくとも１％増加する、請求項１５に記載の方法。
【請求項３０】
核酸配列情報のコンティグをスキャフォールドする方法であって、該方法は：
（ａ）初期の構成を持つ一連のコンティグ配列を得る工程；
（ｂ）一連のペアエンドリードを得る工程；
（ｃ）標準のペアエンドリードの距離度数データを得る工程；
（ｄ）少なくとも１つのペアエンドリードに共存する配列を共有するコンティグのペアをグループ分けする工程であって、それによってグループ分けされたコンティグを生成する工程；及び、
（ｅ）コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データが、初期の構成におけるグループ分けされたコンティグのリードペア度数データに比べて少なくとも５％、標準のペアエンドリードの距離度数データとより近くに近似するように、グループ分けされたコンティグをスキャフォールドする工程、
を含む、方法。
【請求項３１】
スキャフォールドする工程が、グループ分けされたコンティグを順序付けする工程、グループ分けされたコンティグを配向する工程、少なくとも２つのコンティグの端と端を融合する工程、第２のコンティグに１つのコンティグを挿入する工程、および、少なくとも２つの成分のコンティグへ、コンティグを切断する工程、を含む、請求項３０に記載の方法。
【請求項３２】
リードペア距離尤度が、初期の構成におけるグループ分けされたコンティグのリードペア度数データに比べて少なくとも５％増加する時、コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データが、ペアエンドリードの距離度数データとより近くに近似する、請求項３０に記載の方法。
【請求項３３】
リードペア距離尤度は最大化される、請求項３２に記載の方法。
【請求項３４】
コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データと標準的なペアエンドリード距離度数データとの間の差の統計的尺度が、初期の構成におけるグループ分けされたコンティグのリードペア度数データに比べて少なくとも５％減少する時、コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データが、標準的なペアエンドリードの距離度数データとより近くに近似する、請求項３０に記載の方法。
【請求項３５】
コンティグを分離するためにマッピングされるリードペアに関するリードペアの距離度数データと標準的なペアエンドリード距離度数データとの間の距離の統計的尺度は、ＡＮＯＶＡ、ｔ－検定、及びＸ二乗検定の少なくとも１つを含む、請求項３４に記載の方法。
【請求項３６】
コンティグを分離するためにマッピングされるリードペアのリードペア距離が、標準的なペアエンドリード距離度数と比較して得られた順序付けされたコンティグ中のリードペア距離分布の偏差が減少する時、ペアエンドリード距離度数データとより近く一致する、請求項３５に記載の方法。
【請求項３７】
標準的なペアエンドリード距離度数と比較して得られた順序付けされたコンティグ中のリードペア距離分布の偏差が最小化される、請求項３６に記載の方法。
【請求項３８】

Ｎ５０値は、初期の構成におけるグループ分けされたコンティグに比べて少なくとも１％増加する、請求項３０に記載の方法。