



- (51) 국제특허분류:
C12Q 1/68 (2006.01) G06F 19/22 (2011.01)
- (21) 국제출원번호: PCT/KR2013/007288
- (22) 국제출원일: 2013년 8월 13일 (13.08.2013)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2012-0120447 2012년 10월 29일 (29.10.2012) KR
- (71) 출원인: 삼성에스디에스 주식회사 (SAMSUNG SDS CO., LTD.) [KR/KR]; 135-918 서울시 강남구 역삼2동 707-19 일옥빌딩, Seoul (KR).
- (72) 발명자: 박민서 (PARK, Minseo); 135-284 서울시 강남구 대치4동 931-6, Seoul (KR).
- (74) 대리인: 노준태 (NOH, Jun Tae) 등; 137-953 서울시 서초구 서초중앙로 53 6층 (서초동, 대림빌딩), Seoul (KR).

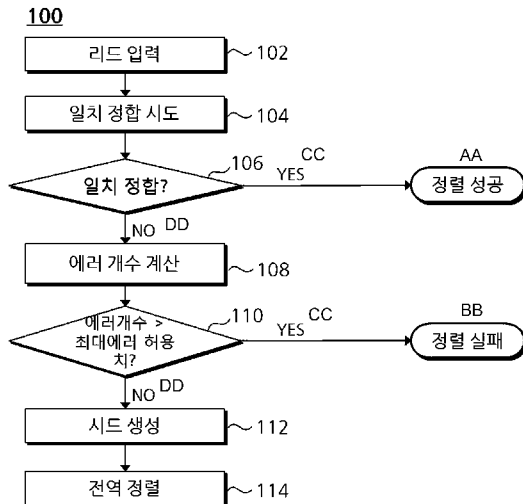
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[다음 쪽 계속]

(54) Title: BASE SEQUENCE RECOMBINATION SYSTEM AND METHOD

(54) 발명의 명칭: 염기 서열 재조합 시스템 및 방법



(57) Abstract: Disclosed are a base sequence recombination system and method. According to one embodiment of the present invention, the base sequence recombination system comprises: a mapping position calculating unit which selects one seed from among a plurality of seeds produced from a read, and calculates a mapping position of the selected seed in a target sequence; and a global alignment unit which calculates a duplication assessment region for the selected seed from the calculated mapping position, assesses whether global alignment has already been carried out within the calculated duplication assessment region, and, if same has not been carried out, carries out global alignment of the selected read in the calculated mapping position.

(57) 요약서: 염기 서열 재조합 시스템 및 방법이 개시된다. 본 발명의 일 실시예에 따른 염기 서열 재조합 시스템은, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 맵핑 위치 계산부; 및 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하고, 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하여, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 전역 정렬부를 포함한다.

102 ... Input read
104 ... Attempt exact matching
106 ... Exact match?
108 ... Calculate number of errors
110 ... Number of errors > Maximum permitted error value?
112 ... Produce seed
114 ... Globally align
AA ... Alignment successful
BB ... Alignment failed
CC ... YES
DD ... NO

공개:

— 국제조사보고서와 함께 (조약 제 21 조(3))

명세서

발명의 명칭: 염기 서열 재조합 시스템 및 방법

기술분야

- [1] 본 발명의 실시예들은 유전체의 염기 서열을 분석하기 위한 기술과 관련된다.

배경기술

- [2] 저렴한 비용과 빠른 데이터 생산으로 인해 대용량의 짧은 서열을 생산하는 차세대 시퀀싱(NGS; Next Generation Sequencing)이 전통적인 생거(Sanger) 시퀀싱 방식을 빠르게 대체하고 있다. 또한 다양한 NGS 서열재조합 프로그램들이 정확도에 초점을 맞추어 개발되었다. 그러나, 최근 차세대 시퀀싱 기술이 발전함에 따라 단편서열을 만들어 내는 비용이 예전의 절반 이하가 되었고, 이에 따라 사용할 수 있는 데이터의 양이 많아지게 되어서, 대용량의 짧은 서열들을 빠른 시간에 정확하게 처리하기 위한 기술이 필요하게 되었다.
- [3] 서열 재조합의 첫 번째 단계는 염기 서열 정렬(alignment) 알고리즘을 통해 리드를 참조 서열의 정확한 위치에 맵핑(mapping)하는 것이다. 여기서의 문제점은 같은 종의 개체라 할지라도 다양한 유전적 변이로 인해 유전체 서열에 차이가 있을 수 있다는 점이다. 또한 시퀀싱 과정에서의 오류로 인해서도 염기 서열에 차이가 생길 수 있다. 따라서 염기 서열 정렬 알고리즘은 이러한 차이와 변이를 효과적으로 고려해서 맵핑 정확도를 높이지 않으면 안 된다.
- [4] 결론적으로, 유전체 정보의 분석을 진행하기 위해서는, 될 수 있는 한 많은 수의 정확한 전체 유전체 정보 데이터가 필요하다. 또 이를 위해서는 무엇보다도 뛰어난 정확도와 큰 처리량을 갖는 염기 서열 정렬 알고리즘을 개발하는 것이 선행되어야 한다. 그러나 종래의 방법들은 이러한 요구 조건들을 만족시키는 데 한계가 있었다.

발명의 상세한 설명

기술적 과제

- [5] 본 발명의 실시예들은 맵핑 정확도를 보장하는 동시에 맵핑 시의 복잡도를 개선하여 처리 속도를 높일 수 있는 염기 서열 재조합 수단을 제공하는 데 그 목적이 있다.

과제 해결 수단

- [6] 상기 과제를 해결하기 위한 본 발명의 일 실시예에 따른 염기 서열 재조합 시스템은, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 맵핑 위치 계산부; 및 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하고, 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하여, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 전역 정렬부를 포함한다.

- [7] 한편, 상기 과제를 해결하기 위한 본 발명의 일 실시예에 따른 염기 서열 재조합 방법은, 맵핑 위치 계산부에서, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 단계; 전역 정렬부에서, 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하는 단계; 및 상기 전역 정렬부에서, 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하여, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 단계를 포함한다.
- [8] 한편, 본 발명의 일 실시예에 따른 장치는 하나 이상의 프로세서; 메모리; 및 하나 이상의 프로그램을 포함하는 장치로서, 상기 하나 이상의 프로그램은 상기 메모리에 저장되고 상기 하나 이상의 프로세서에 의해 실행되도록 구성되며, 상기 프로그램은, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 단계; 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하는 단계; 및 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하여, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 단계를 실행하기 위한 명령어들을 포함한다.

발명의 효과

- [9] 본 발명의 실시예들에 따른 경우, 염기 서열 정렬 시 전역 정렬을 수행한 위치를 기억해 두었다가 해당 위치의 주변에서는 전역 정렬을 수행하지 않으므로, 염기 서열 정렬 과정에서 많은 시간이 소요되는 전역 정렬의 수행 횟수를 감소시킬 수 있으며, 이에 따라 염기 서열 정렬 시간을 대폭 감소시킬 수 있다.
- [10] 또한 상기와 같이 전역 정렬을 중복하여 수행하지 않는 중복 영역의 크기를 리드의 길이에 비례하도록 설정함으로써, 염기 서열 정렬 시간을 감소하면서도 염기 서열 정렬의 정확도를 유지할 수 있는 장점이 있다.

도면의 간단한 설명

- [11] 도 1은 본 발명의 일 실시예에 따른 염기 서열 재조합 방법을 설명하기 위한 도면이다.
- [12] 도 2는 본 발명의 일 실시예에 따른 염기 서열 정렬 방법의 에러 개수 계산 과정을 예시하기 위한 도면이다.
- [13] 도 3은 본 발명의 일 실시예에 따른 전역 정렬 과정을 설명하기 위한 순서도이다.
- [14] 도 4a 내지 4e는 본 발명의 일 실시예에 따른 전역 정렬 과정을 예를 들어 설명하기 위한 도면이다.
- [15] 도 5는 본 발명의 일 실시예에 따른 염기 서열 재조합 시스템을 도시한 블록도이다.

발명의 실시를 위한 형태

- [16] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.
- [17] 본 발명을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [18] 본 발명의 기술적 사상은 청구범위에 의해 결정되며, 이하의 실시예는 본 발명의 기술적 사상을 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 효율적으로 설명하기 위한 일 수단일 뿐이다.
- [19]
- [20] 본 발명의 실시예들을 상세히 설명하기 앞서, 먼저 본 발명에서 사용되는 용어들에 대하여 설명하면 다음과 같다.
- [21] 먼저, "리드(read)"란 게놈 시퀀서(genome sequencer)에서 출력되는 짧은 길이의 염기서열 데이터이다. 리드의 길이는 게놈 시퀀서의 종류에 따라 일반적으로 35~500bp(base pair) 정도로 다양하게 구성되며, 일반적으로 DNA 염기의 경우 A, C, G, T의 알파벳 문자로 표현된다.
- [22] "대상 염기 서열"이란 상기 리드들로부터 전체 염기 서열을 생성하는 데 참조가 되는 염기 서열(reference sequence)을 의미한다. 염기 서열 분석에서는 게놈 시퀀서에서 출력되는 다량의 리드들을 대상 염기 서열에 맵핑함으로써 전체 염기 서열을 완성하게 된다. 본 발명에서 상기 대상 염기 서열은 염기 서열 분석 시 미리 설정된 서열(예를 들어 인간의 전체 염기 서열 등)일 수도 있으며, 또는 게놈 시퀀서에서 만들어진 염기 서열을 대상 염기 서열로 사용할 수도 있다.
- [23] "베이스(base)"는 대상 염기 서열 및 리드를 구성하는 최소 단위이다. 전술한 바와 같이 DNA 염기의 경우 A, C, G 및 T의 네 종류의 알파벳 문자로 구성될 수 있으며, 이들 각각을 베이스라 표현한다. 다시 말해 DNA 염기의 경우 4개의 베이스로 표현되며, 이는 리드 또한 마찬가지이다.
- [24] "시드(seed)"란 리드의 맵핑을 위하여 리드와 대상 염기 서열을 비교할 때의 단위가 되는 시퀀스이다. 이론적으로 리드를 대상 염기 서열에 맵핑하기 위해서는 리드 전체를 대상 염기 서열의 가장 첫 부분부터 순차적으로 비교해 나가면서 리드의 맵핑 위치를 계산하여야 한다. 그러나 이와 같은 방법의 경우 하나의 리드를 맵핑하는 데 너무 많은 시간 및 컴퓨팅 파워가 요구되므로, 실제로는 리드의 일부분으로 구성된 조각인 시드를 먼저 대상 염기 서열에 맵핑함으로써 전체 리드의 맵핑 후보 위치를 찾아 내고 해당 후보 위치에 전체

리드를 맵핑(전역 정렬; Global Alignment)하게 된다.

[25]

[26] 도 1은 본 발명의 일 실시예에 따른 염기 서열 재조합 방법(100)을 설명하기 위한 도면이다. 본 발명의 실시예에서, 염기 서열 재조합 방법(100)이란 게놈 시퀀서(genome)에서 출력되는 리드를 대상 염기 서열과 비교하여 리드의 상기 대상 염기 서열에서의 맵핑(또는 정렬) 위치를 결정하여 전체 서열을 완성하는 일련의 과정을 의미한다.

[27] 먼저, 게놈 시퀀서(genome sequencer)로부터 리드가 입력되면(102), 리드 전체와 상기 대상 염기 서열과의 일치 정합(exact matching)을 시도한다(104). 만약 상기 시도 결과 리드 전체에 대한 일치 정합이 성공한 경우에는 이후의 정렬 단계를 수행하지 않고 정렬에 성공한 것으로 판단한다(106).

[28] 인간의 염기 서열을 대상으로 한 실험 결과, 게놈 시퀀서에서 출력되는 100만 개의 리드를 인간의 염기 서열에 일치 정합할 경우, 각 리드의 길이가 755bp일 때, 총 200만회의 정렬 중(정방향 시퀀스 100만회, 역상보(reverse complement) 방향 시퀀스 100만회) 231,564회의 일치 정합이 발생하는 것으로 나타났다. 따라서 상기 104 단계의 수행 결과 약 11.6%만큼의 정렬 소요량을 감소시킬 수 있었다.

[29] 그러나, 이와 달리 상기 106 단계에서 해당 리드가 일치 정합되지 않는 것으로 판단되는 경우에는 해당 리드를 상기 대상 서열에 정렬했을 때 나타날 수 있는 에러 개수(mismatch)를 계산한다(108).

[30] 도 2는 상기 108 단계에서의 상기 에러 개수 계산 과정을 예시하기 위한 도면이다. 먼저, 도 2의 (a)에 도시된 바와 같이 에러 개수의 최초값을 0으로 설정하고($mismatch = 0$) 리드의 가장 첫 번째 베이스부터 오른쪽으로 한 베이스씩 이동하면서 일치 정합을 시도한다. 이때 (b)에 도시된 바와 같이 리드의 특정 베이스(화살표로 표시)에서부터 더 이상 일치 정합이 불가능하다고 가정하자. 이 경우는 리드의 정합 시작 위치부터 현재 위치 사이의 구간 어딘가에서 에러가 발생한 것을 의미한다. 따라서 이 경우에는 에러 개수의 값을 1만큼 증가시키고($mismatch = 0 \rightarrow 1$), 다음 위치에서 새로 일치 정합을 시작한다(도면에서 (c)로 표기). 이후 또 다시 일치 정합이 불가능하다고 판단되는 경우에는, 일치 정합을 새로 시작한 위치부터 현재 위치 사이의 구간 어디에서 다시 에러가 발생한 것이므로, 에러 개수 값을 다시 1만큼 증가시키고($mismatch = 1 \rightarrow 2$), 다음 위치에서 새로 일치 정합을 시작한다(도면에서 (d)로 표기). 이와 같은 과정을 거쳐 리드의 끝까지 도달한 경우의 에러 개수 값이 해당 리드의 발생할 수 있는 에러 개수 값이 된다. 즉, 도시된 실시예에서 리드의 에러 개수는 2가 된다.

[31] 상기와 같은 과정을 거쳐 리드의 에러 개수가 계산되면, 계산된 에러 개수 값이 기 설정된 최대 에러 허용치(maxError)를 초과하는지의 여부를 판단하고(110), 초과하는 경우 해당 리드에 대한 정렬이 실패한 것으로 판단하여 정렬을

종료한다.

- [32] 전술한 인간의 염기 서열을 대상으로 한 실험에서, 최대 에러 허용치(maxError)를 3으로 하고 나머지 리드들의 에러 개수를 계산한 결과, 총 844,891회에 해당하는 리드들이 상기 최대 에러 허용치를 초과하는 것으로 나타났다. 즉, 상기 108 단계의 수행 결과 약 42.2%만큼의 정렬 소요량을 감소시킬 수 있었다.
- [33] 그러나 이와 달리 상기 110 단계에서의 판단 결과, 계산된 에러 개수 값이 상기 최대 에러 허용치 이하인 경우에는 다음과 같은 과정을 거쳐 해당 리드에 대한 정렬을 수행한다.
- [34] 먼저, 상기 리드로부터 복수 개의 시드를 생성하고(112), 생성된 상기 복수 개의 시드를 이용하여 상기 리드에 대한 전역 정렬(global alignment)을 수행한다(114). 이때 상기 전역 정렬의 결과 리드의 에러 개수가 기 설정된 최대 에러 허용치(maxError)를 초과하는 경우에는 정렬 실패로, 그렇지 않은 경우에는 정렬에 성공한 것으로 판단된다.
- [35] 이하에서는 상기 112 단계 및 114 단계의 구체적인 과정을 상세히 설명한다.
- [36]
- [37] 리드로부터 복수 개의 시드 생성(112)
- [38]
- [39] 본 단계는 본격적으로 리드의 정렬을 수행하기 위하여 리드로부터 복수 개의 작은 조각인 시드를 생성하는 단계이다. 본 단계에서는 상기 리드의 일부 또는 전체를 고려하여 복수 개의 시드들을 생성하게 된다. 예를 들어, 리드의 전체, 또는 특정 구간을 복수 개의 조각으로 분할하거나, 분할된 조각들을 조합함으로써 시드들을 생성할 수 있다. 이 경우 생성된 시드들은 서로 연속적으로 연결될 수 있으나 반드시 그러한 것은 아니며, 리드 내에서 서로 떨어진 조각들의 조합으로 시드들을 구성하는 것 또한 가능하다. 또한, 생성되는 시드들이 반드시 동일한 길이를 가질 필요는 없으며, 하나의 리드 내에서 다양한 길이를 가지는 시드들을 생성하는 것 또한 가능하다. 요컨대, 본 발명에서 리드로부터 시드를 생성하는 방법은 특별히 제한되지 않으며, 리드의 일부 또는 전체로부터 시드를 추출하는 다양한 알고리즘이 제한 없이 사용될 수 있다.
- [40]
- [41] 전역 정렬(Global Alignment)(114)
- [42]
- [43] 상기와 같은 과정을 거쳐 시드들이 생성되면, 다음으로 생성된 시드들을 이용하여 상기 리드의 상기 대상 서열에 대한 전역 정렬을 수행한다. 구체적으로, 본 단계에서는 상기 112 단계에서 생성된 시드들을 이용하여, 각 시드들의 상기 대상 서열에서의 맵핑 위치에서 차례로 전역 정렬을 수행함으로써 상기 리드를 상기 대상 서열에 맵핑하게 된다.
- [44] 도 3은 본 발명의 일 실시예에 따른 전역 정렬 과정(114)을 설명하기 위한

순서도이다. 먼저, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고(302), 선택된 시드의 대상 서열에서의 맵핑 위치를 계산한다(304). 본 발명의 실시예에서, 특별한 한정 없이 단순히 시드의 "맵핑 위치"라고만 기재할 경우에는 해당 시드의 첫 번째 베이스에 대응하는 대상 서열의 위치를 의미하며, 시드의 "k번째 맵핑 위치"는 해당 시드의 k 번째 베이스에 대응하는 대상 서열의 위치를 의미한다.

[45] 다음으로, 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산한다(306). 상기 중복 판단 영역은, 예를 들어 상기 선택된 시드의 상기 대상 서열에서의 k 번째 맵핑 위치($1 \leq k \leq N$, N은 상기 선택된 시드의 길이)로부터의 거리차가 설정된 기준값 이내인 영역으로 설정될 수 있다.

[46] 또는, 상기 중복 판단 영역은, 다음의 수학적 식 1에 의하여 계산될 수도 있다.

[47]

[48] [수학적 식 1]

[49] $m_a - V \leq \text{중복 판단 영역} \leq m_b + V$

[50] (이때, m_a 는 상기 선택된 시드의 a번째 맵핑 위치($1 \leq a \leq N$), m_b 는 상기 선택된 시드의 b번째 맵핑 위치($1 \leq b \leq N$), N은 상기 선택된 시드의 길이, V는 기준값)

[51]

[52] 상기와 같은 방법으로 중복 판단 영역이 계산되면, 다음으로 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단한다(308). 이때, 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부는, 이전 단계의 전역 정렬시의 맵핑 위치가 (즉, 전역 정렬을 수행한 시드의 첫 번째 맵핑 위치가) 상기 중복 판단 영역에 포함되는지의 여부로 판단할 수 있다. 만약 상기 판단 결과 상기 중복 판단 영역 내에서 전역 정렬이 수행된 경우 상기 302 단계에서 선택된 시드에 대한 전역 정렬은 수행되지 않으며, 이 경우 생성된 상기 시드들 중 아직 전역 정렬을 수행하지 않고 남아 있는 시드가 존재하는지의 여부를 판단하고(314), 존재하는 경우 상기 302 단계로 돌아가 남은 시드 중 새롭게 선택된 시드에 대하여 상기 과정을 반복하게 된다. 이때 만약 상기 314 단계의 판단 결과 남은 시드가 존재하지 않는 경우에는 상기 리드에 대한 정렬은 실패한 것으로 판단된다.

[53] 한편, 상기 308 단계의 판단 결과 해당 영역에서 전역 정렬이 수행되지 않은 경우에는 계산된 상기 맵핑 위치에서 상기 리드에 대한 전역 정렬을 수행하여(310), 계산된 에러의 개수가 기 설정된 최대 에러 허용치를 초과하는지의 여부를 판단한다(312). 만약 상기 312 단계의 판단 결과 해당 맵핑 위치에서의 에러 개수가 상기 최대 에러 허용치 이내인 경우에는 리드의 정렬에 성공한 것으로 판단된다. 그러나 만약 상기 에러 개수가 최대 에러 허용치를 초과하는 경우에는, 다음으로 남은 시드가 존재하는지를 판단하고(314). 존재하는 경우 상기 302 단계로 돌아가 남은 시드 중 새롭게 선택된 시드에 대하여 상기 과정을 반복하게 된다. 이때 만약 상기 314 단계의 판단 결과 남은

시드가 존재하지 않는 경우에는 상기 리드에 대한 정렬은 실패한 것으로 판단된다.

- [54] 전술한 306 단계 및 308 단계를 좀 더 상세히 설명하면 도 4a 내지 4e와 같다. 도시된 실시예에서, 리드로부터 3개의 시드(SEED 1, SEED 2, SEED 3)가 추출되었으며, 각각의 대상 염기 서열에서의 맵핑 위치는 각각 2001bp, 2101bp, 및 2301bp, 각 시드의 전역 정렬 여부를 판단하기 위한 기준값은 128bp, 각 시드의 길이는 30bp, 리드의 정렬을 위하여 SEED 1, SEED 2 및 SEED 3의 순서로 전역 정렬을 수행한다고 가정하자. 먼저 SEED 1의 경우 이전에 수행된 전역 정렬이 존재하지 않으므로 정상적으로 해당 위치(2001bp)에서 리드를 대상 서열에 전역 정렬한다. 그러나 다음에 맵핑될 SEED 2의 경우에는 SEED 2의 맵핑 위치로부터 계산되는 중복 판단 영역에 따라 전역 정렬 수행 여부가 달라진다.
- [55] 먼저, 도 4a에 도시된 바와 같이, 상기 중복 판단 영역은 시드의 첫 번째 맵핑 위치로부터의 거리차가 기준값 만큼 떨어진 영역 내로 정의될 수 있다. 즉, 도시된 실시예에서 SEED 2의 중복 판단 영역은 SEED 2의 첫 번째 맵핑 위치인 2101bp의 앞뒤로 128bp에 해당하는 영역이 된다(도면에서 회색으로 표시된 영역). 이 경우에는 상기 중복 판단 영역 내에서 SEED 1에 대한 전역 정렬이 수행되었으므로, SEED 2의 맵핑 위치에서의 전역 정렬은 수행되지 않는다.
- [56] 다음으로, 도 4b에 도시된 바와 같이, 상기 중복 판단 영역은 시드의 마지막 맵핑 위치로부터의 거리차가 기준값 만큼 떨어진 영역 내로 정의될 수 있다. 즉, 도시된 실시예에서 SEED 2의 중복 판단 영역은 SEED 2의 마지막 맵핑 위치인 2130bp의 앞뒤로 128bp에 해당하는 영역이 된다(도면에서 회색으로 표시된 영역). 이 경우에는 기 전역 정렬이 수행된 SEED 1의 맵핑 위치(2001bp)가 상기 중복 판단 영역의 밖에 있으므로, SEED 2의 맵핑 위치에서 전역 정렬이 수행된다.
- [57] 도 4c는 상기 도 4a 및 4b의 실시예를 일반화하여 시드의 k 번째 맵핑 위치로부터($1 \leq k \leq N$, 이때 N 은 시드의 길이)의 거리차가 기준값 만큼 떨어진 영역으로 중복 판단 영역을 설정하는 실시예를 나타낸 것이다. 이 경우는 상기 k 값에 따라 SEED 2의 전역 정렬 여부가 달라지게 된다.
- [58] 한편, 도 4d에 도시된 바와 같이, 상기 중복 판단 영역은 시드의 첫 번째 맵핑 위치로부터 대상 서열의 앞쪽 방향으로 기준값 만큼 떨어진 위치로부터 시드의 마지막 맵핑 위치로부터 대상 서열의 뒤쪽 방향으로 기준값 만큼 떨어진 위치까지를 포함하도록 구성될 수도 있다. 즉, 이 경우는 상기 도 4a 및 4b에서의 중복 판단 영역을 합친 것과 동일한 결과가 된다. 도 4e는 이를 일반화하여 전술한 수학식 1에 따라 중복 판단 영역을 설정하는 실시예를 나타낸 것이다.
- [59] 이상과 같이 하나의 시드에 대한 전역 정렬이 수행된 경우 그 주변에 대한 전역 정렬을 수행하지 않는 이유는 다음과 같다. 전역 정렬의 후보가 되는 각각의 시드들은 하나의 리드에서 나온 것이므로, 각 시드들이 대상 염기 서열에서

유사한 구간 내에 맵핑된다는 것은 결국 해당 리드는 해당 구간 내에 맵핑될 가능성이 매우 높다는 것을 의미한다. 따라서 이 경우에는 해당 구간 내의 맵핑되는 여러 개의 시드 중 하나의 시드에 대해서만 전역 정렬을 수행하더라도 해당 위치에 리드를 맵핑하는 것이 충분히 가능하다. 또한, 이를 반대로 판단해보면, 비슷한 구간 내에 맵핑되는 여러 개의 시드 중 하나의 시드에 대한 전역 정렬 결과 리드가 맵핑되지 않는 경우에는 다른 시드 또한 해당 구간에는 맵핑되지 않을 가능성이 높다는 것을 의미한다. 따라서 본 발명의 실시예에서는 각 시드들에 대하여 중복 판단 영역을 설정하고, 해당 영역 내에서 전역 정렬이 수행된 경우에는 전역 정렬을 중복하여 수행하지 않음으로써 매우 많은 시간이 소요되는 전역 정렬의 횟수를 효과적으로 감소시킬 수 있도록 구성하였다. 구체적으로, 본 발명의 전역 정렬 방법을 사용한 알고리즘과 사용하지 않은 알고리즘 사이에는 약 30~35배의 속도 차이가 발생하는 것으로 나타났다.

- [60] 한편, 상기 기준값은 상기 리드의 길이에 비례하도록 설정될 수 있으며, 구체적으로 상기 기준값은 상기 리드 길이의 100% 내지 170%로 설정될 수 있다. 상기 기준값이 리드의 길이에 비례하는 이유는, 전역 정렬이 리드를 이용하여 수행되기 때문이다. 즉, 맵핑 위치에서부터 리드의 길이만큼의 구간은 이미 전역 정렬이 수행되었으므로 중복하여 전역 정렬을 수행할 필요가 없기 때문이다. 또한, 상기 기준값이 리드 길이의 170%까지 확장되는 이유는 염기 서열의 추가(insertion) 또는 삭제(deletion) 등으로 인하여 리드 또는 대상 염기 서열에 오차가 발생할 수 있으므로, 이를 감안한 것이다. 이와 같이 상기 기준값을 리드의 길이에 연동하여 변화하도록 할 경우 전술한 바와 같이 염기 서열 재조합 알고리즘의 속도를 향상시키면서도 맵핑의 정확도를 유지할 수 있는 장점이 있다.
- [61]
- [62] 도 5는 본 발명의 일 실시예에 따른 염기 서열 재조합 시스템(500)의 블록도이다. 본 발명의 일 실시예에 따른 염기 서열 재조합 시스템(500)은 전술한 염기 서열 재조합 방법을 수행하기 위한 장치로서, 시드 생성부(502), 맵핑 위치 계산부(504) 및 전역 정렬부(506)를 포함한다.
- [63] 시드 생성부(502)는 게놈 시퀀서로부터 얻어진 리드로부터 복수 개의 시드들을 생성한다. 전술한 바와 같이, 시드 생성부(502)에서 리드로부터 시드를 생성하는 방법은 특별히 제한되지 않으며, 리드의 일부 또는 전체로부터 시드를 추출하는 다양한 알고리즘이 제한 없이 사용될 수 있다.
- [64] 맵핑 위치 계산부(504)는, 시드 생성부(502)에서 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드에 대한 대상 서열에서의 맵핑 위치를 계산한다.
- [65] 전역 정렬부(506)는 맵핑 위치 계산부(504)에서 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하고, 계산된 상기 중복 판단

영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하여, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 리드에 대한 전역 정렬(Global Alignment)을 수행한다. 이 경우 상기 중복 판단 영역의 계산과 관련된 상세한 사항은 전술하였으므로 여기서는 그 상세한 설명을 생략한다.

[66]

[67] 한편, 본 발명의 실시예는 본 명세서에서 기술한 방법들을 컴퓨터상에서 수행하기 위한 프로그램을 포함하는 컴퓨터 판독 가능 기록매체를 포함할 수 있다. 상기 컴퓨터 판독 가능 기록매체는 프로그램 명령, 로컬 데이터 파일, 로컬 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야에서 통상의 지식을 가진 자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광 기록 매체, 플로피 디스크와 같은 자기-광 매체, 및 롬, 램, 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

[68] 이상에서 대표적인 실시예를 통하여 본 발명에 대하여 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다.

[69] 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

[70]

[71] <부호의 설명>

[72] 500: 염기 서열 재조합 시스템

[73] 502: 시드 생성부

[74] 504: 맵핑 위치 계산부

[75] 506: 전역 정렬부

청구범위

- [청구항 1] 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 맵핑 위치 계산부; 및
계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하고, 계산된 상기 중복 판단 영역 내에서 전역 정렬이 수행되었는지의 여부를 판단하며, 상기 판단 결과 상기 전역 정렬이 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 전역 정렬부를 포함하는 염기 서열 재조합 시스템.
- [청구항 2] 청구항 1에 있어서,
상기 중복 판단 영역은, 상기 선택된 시드의 상기 대상 서열에서의 k 번째 맵핑 위치($1 \leq k \leq N$, N 은 상기 선택된 시드의 길이)로부터의 거리차가 설정된 기준값 이내인 영역인, 염기 서열 재조합 시스템.
- [청구항 3] 청구항 2에 있어서,
상기 기준값은 상기 리드의 길이에 비례하도록 설정되는, 염기 서열 재조합 시스템.
- [청구항 4] 청구항 3에 있어서,
상기 기준값은 상기 리드 길이의 100% 내지 170%로 설정되는, 염기 서열 재조합 시스템.
- [청구항 5] 청구항 1에 있어서,
상기 중복 판단 영역은, 다음의 수학적식

$$m_a - V \leq \text{중복 판단 영역} \leq m_b + V$$
(이때, m_a 는 상기 선택된 시드의 a 번째 맵핑 위치($1 \leq a \leq N$), m_b 는 상기 선택된 시드의 b 번째 맵핑 위치($1 \leq b \leq N$), N 은 상기 선택된 시드의 길이, V 는 기준값)
에 의하여 계산되는, 염기 서열 재조합 시스템.
- [청구항 6] 청구항 5에 있어서,
상기 기준값은 상기 리드의 길이에 비례하도록 설정되는, 염기 서열 재조합 시스템.
- [청구항 7] 청구항 6에 있어서,
상기 기준값은 상기 리드 길이의 100% 내지 170%로 설정되는, 염기 서열 재조합 시스템.
- [청구항 8] 청구항 1에 있어서,
상기 전역 정렬부는, 전역 정렬을 기 수행한 시드의 맵핑 위치가 상기 중복 판단 영역 내에 포함되는 경우, 상기 중복 판단 영역

- 내에서 전역 정렬이 수행된 것으로 판단하는, 염기 서열 재조합 시스템.
- [청구항 9] 맵핑 위치 계산부에서, 리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고, 선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 단계;
전역 정렬부에서, 계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하는 단계; 및
상기 전역 정렬부에서, 계산된 상기 중복 판단 영역 내에서 전역 정렬이 기 수행되었는지의 여부를 판단하 판단하고, 상기 판단 결과 상기 전역 정렬이, 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기 선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는 단계를 포함하는 염기 서열 재조합 방법.
- [청구항 10] 청구항 9에 있어서,
상기 중복 판단 영역은, 상기 선택된 시드의 상기 대상 서열에서의 k 번째 맵핑 위치($1 \leq k \leq N$, N은 상기 선택된 시드의 길이)로부터의 거리차가 설정된 기준값 이내인 영역인, 염기 서열 재조합 방법.
- [청구항 11] 청구항 10에 있어서,
상기 기준값은 상기 리드의 길이에 비례하도록 설정되는, 염기 서열 재조합 방법.
- [청구항 12] 청구항 11에 있어서,
상기 기준값은 상기 리드 길이의 100% 내지 170%로 설정되는, 염기 서열 재조합 방법.
- [청구항 13] 청구항 9에 있어서,
상기 중복 판단 영역은, 다음의 수학적 식

$$m_a - V \leq \text{중복 판단 영역} \leq m_b + V$$
(이때, m_a 는 상기 선택된 시드의 a번째 맵핑 위치($1 \leq a \leq N$), m_b 는 상기 선택된 시드의 b번째 맵핑 위치($1 \leq b \leq N$), N은 상기 선택된 시드의 길이, V는 기준값)
에 의하여 계산되는, 염기 서열 재조합 방법.
- [청구항 14] 청구항 13에 있어서,
상기 기준값은 상기 리드의 길이에 비례하도록 설정되는, 염기 서열 재조합 방법.
- [청구항 15] 청구항 14에 있어서,
상기 기준값은 상기 리드 길이의 100% 내지 170%로 설정되는, 염기 서열 재조합 방법.
- [청구항 16] 청구항 9에 있어서,
상기 전역 정렬부는, 전역 정렬을 기 수행한 시드의 맵핑 위치가

상기 중복 판단 영역 내에 포함되는 경우, 상기 중복 판단 영역 내에서 전역 정렬이 수행된 것으로 판단하는, 염기 서열 재조합 방법.

[청구항 17]

하나 이상의 프로세서;

메모리; 및

하나 이상의 프로그램을 포함하는 장치로서,

상기 하나 이상의 프로그램은 상기 메모리에 저장되고 상기 하나 이상의 프로세서에 의해 실행되도록 구성되며,

상기 프로그램은,

리드로부터 생성된 복수 개의 시드들 중 하나의 시드를 선택하고,

선택된 시드의 대상 서열에서의 맵핑 위치를 계산하는 단계;

계산된 상기 맵핑 위치로부터 상기 선택된 시드에 대한 중복 판단 영역을 계산하는 단계; 및

계산된 상기 중복 판단 영역 내에서 전역 정렬이 기

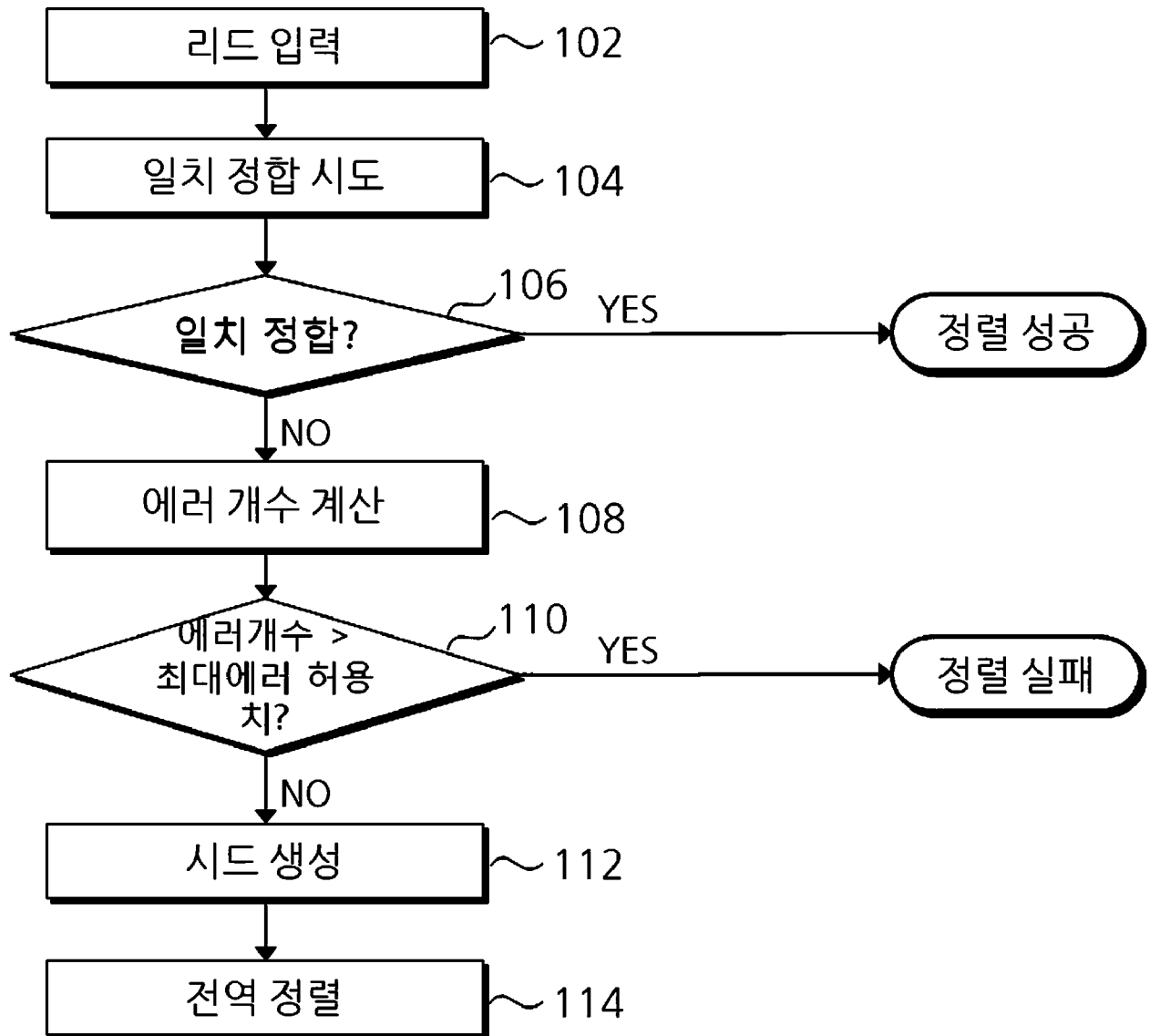
수행되었는지의 여부를 판단하고, 상기 판단 결과 상기 전역

정렬이 수행되지 않은 경우 상기 계산된 맵핑 위치에서 상기

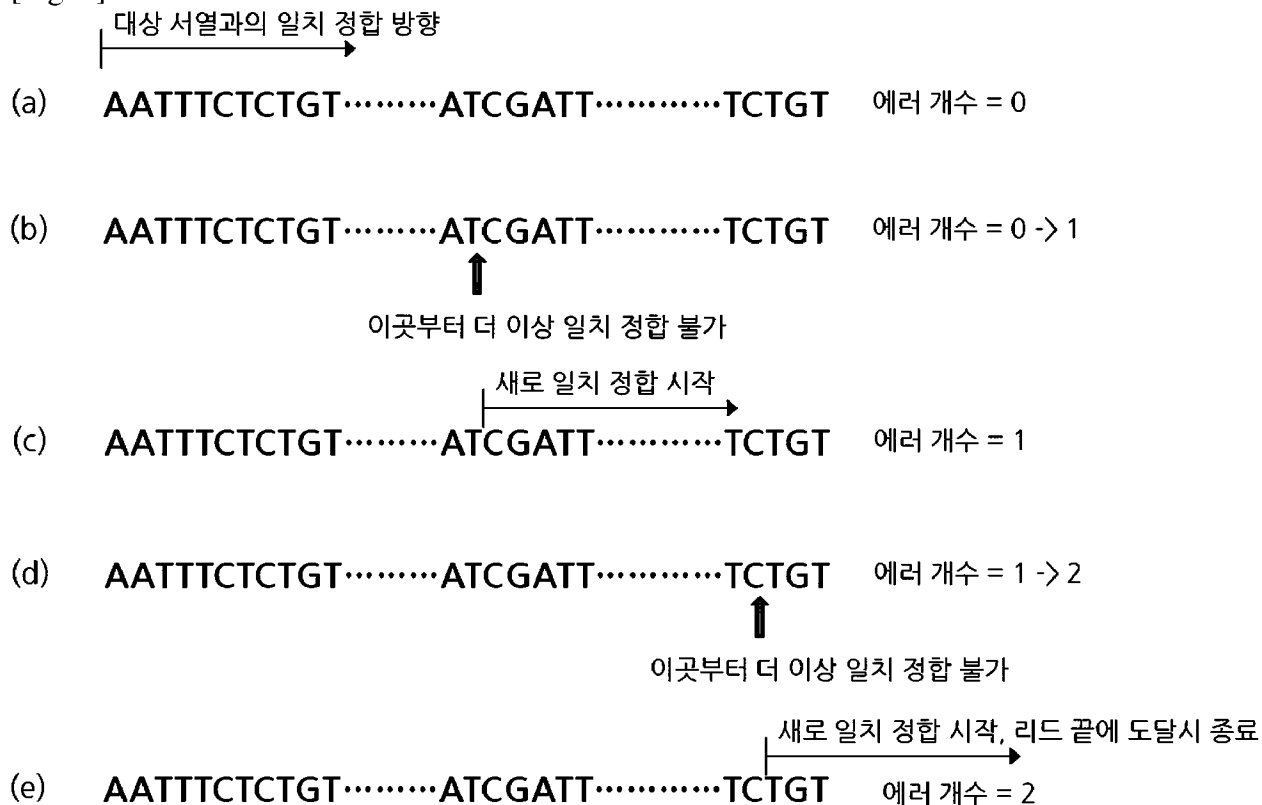
선택된 리드에 대한 전역 정렬(Global Alignment)을 수행하는

단계를 실행하기 위한 명령어들을 포함하는 장치.

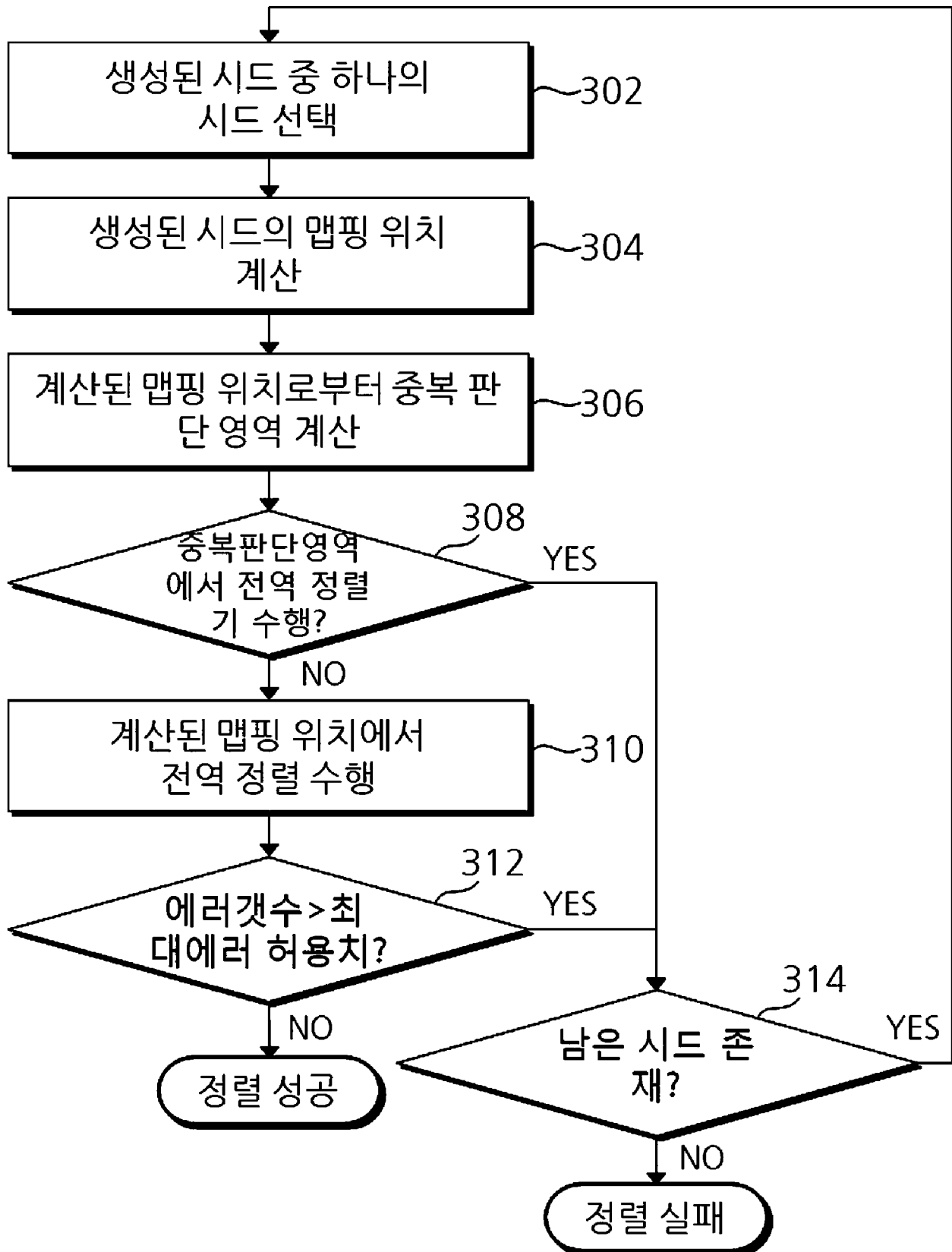
[Fig. 1]

100

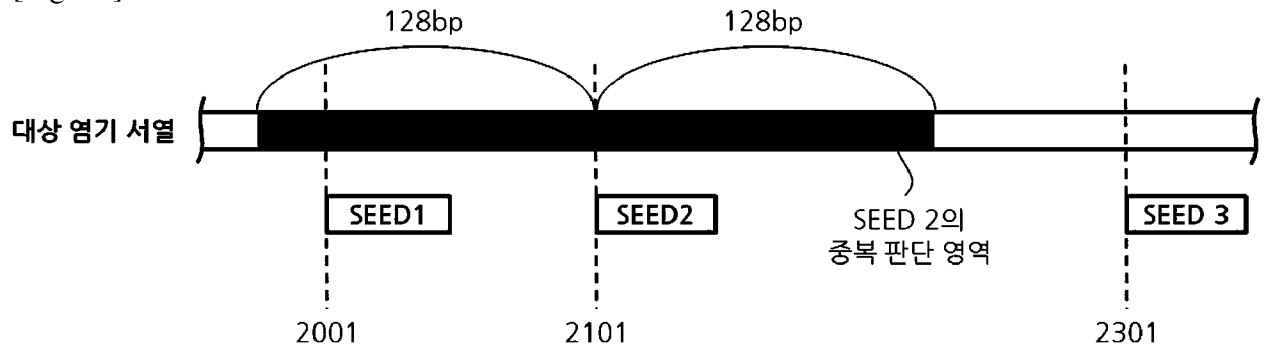
[Fig. 2]



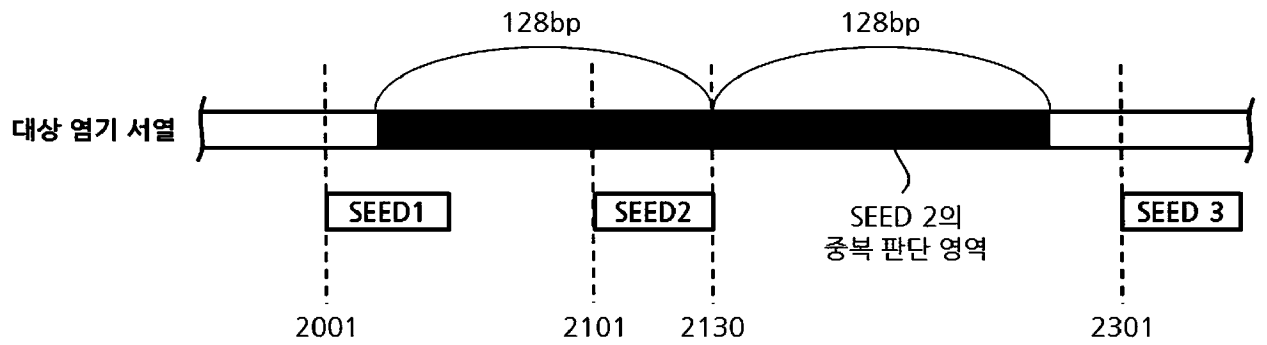
[Fig. 3]

114

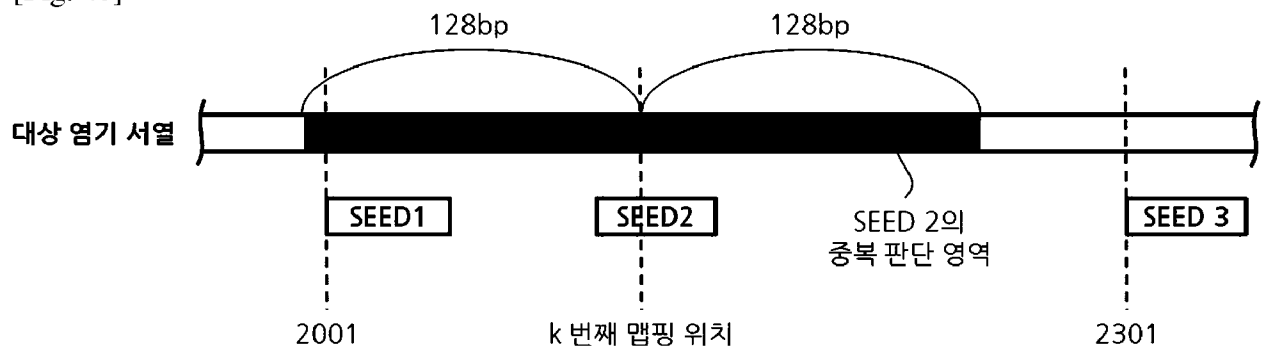
[Fig. 4a]



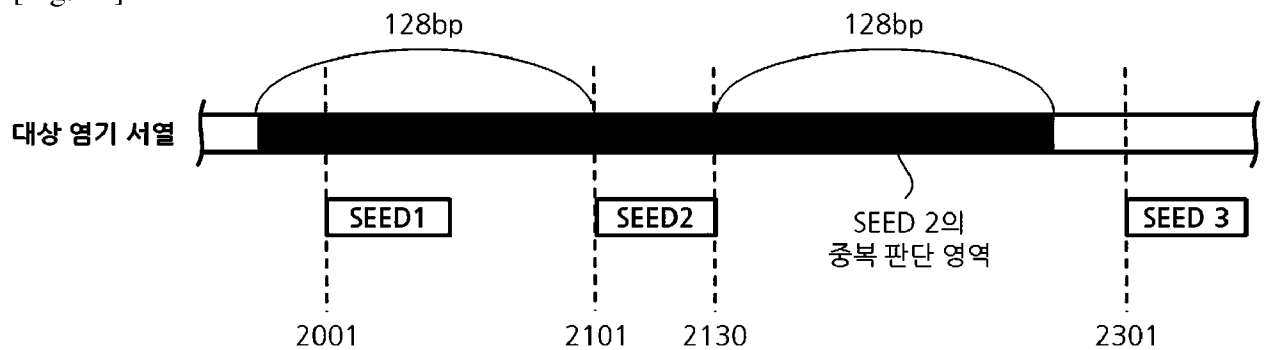
[Fig. 4b]



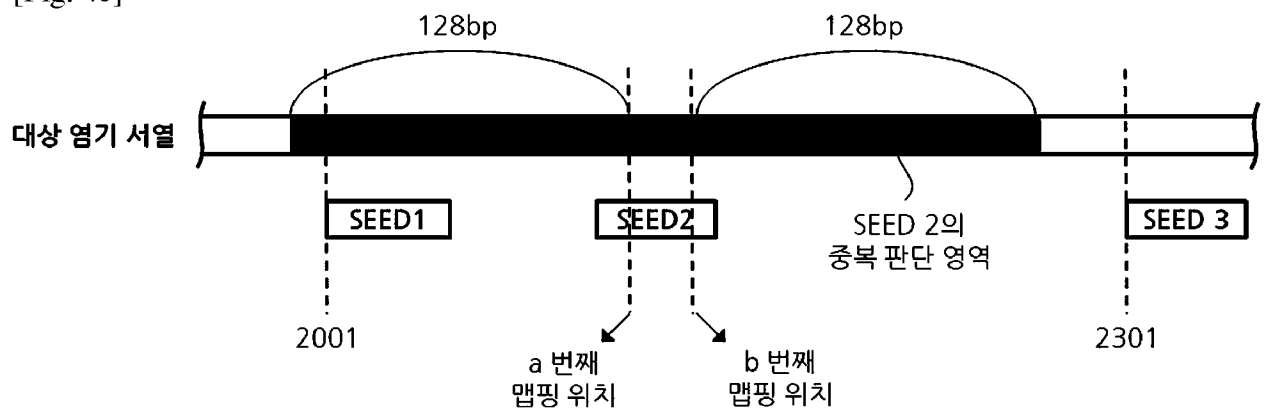
[Fig. 4c]



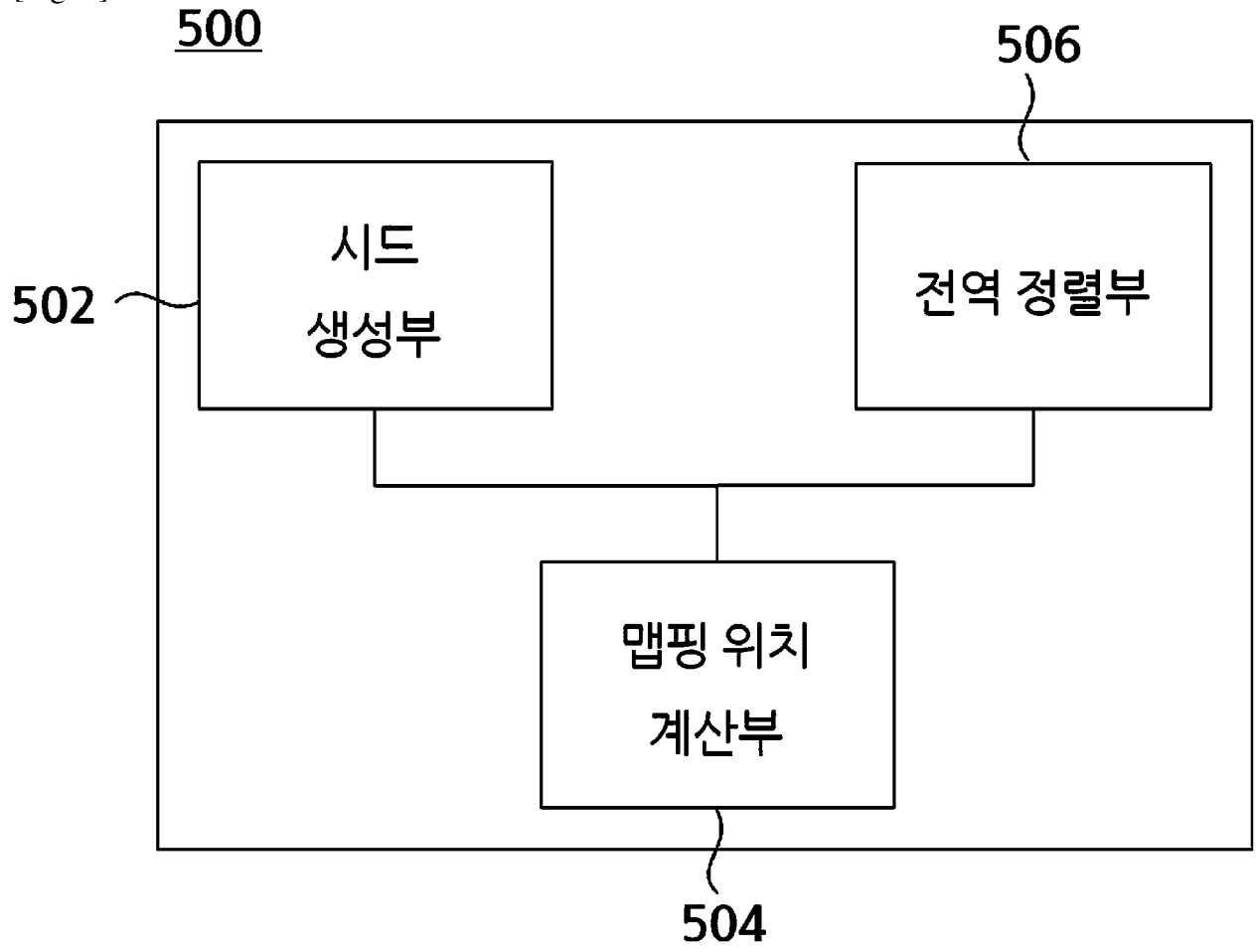
[Fig. 4d]



[Fig. 4e]



[Fig. 5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2013/007288**A. CLASSIFICATION OF SUBJECT MATTER*****C12Q 1/68(2006.01)i. G06F 19/22(2011.01)i***

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q 1/68; G06F 19/00; G06F 19/22

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

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

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

eKOMPASS (KIPO internal) & Keywords: nucleotide sequence recombination, seed, mapping, overlap, global alignment

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PARK, Tae Won et al., "Method for Time Efficient Processing of Next Generation Sequencing Reads Using Hitting Set Problem" Journal of KISSE: Computer Systems and Theory, vol. 38, no. 6, pp. 275-280 (2011) See abstract and pages 277-279.	1-17
A	YEO, Yun-Ku et al., "BARM : A Metagenome Binning Method using Genome Assembly and Reference Alignment" Journal of KISSE: Databases , vol. 38, no. 2, pp. 72-83 (2011) See the entire document.	1-17
A	KR 10-2011-0049005 A (SAMSUNG SDS CO.,LTD.) 12 May 2011 See the entire document.	1-17
A	US 2011-0270533 A1 (ZHANG et al.) 03 November 2011 See the entire document.	1-17
A	PARK, Min Seo et al., "Survey of Short Reads Alignment for Next-Generation Sequencing" Journal of KISSE: Software and Applications, vol. 38, no. 10, pp. 503-511 (2011) See the entire document.	1-17



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

29 OCTOBER 2013 (29.10.2013)

Date of mailing of the international search report

01 NOVEMBER 2013 (01.11.2013)

Name and mailing address of the ISA/KR



Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2013/007288

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2011-0049005 A	12/05/2011	KR 10-1201626 B1	08/11/2012
US 2011-0270533 A1	03/11/2011	WO 2011-137368 A2	03/11/2011
		WO 2011-137368 A3	16/02/2012

A. 발명이 속하는 기술분류(국제특허분류(IPC)) C12Q 1/68(2006.01)i, G06F 19/22(2011.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) C12Q 1/68; G06F 19/00; G06F 19/22 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 염기서열 재조합, 시드, 맵핑, 중복, 전역정렬		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	박태원 외, 'Hitting Set 문제를 이용한 차세대 염기서열 분석 리드의 처리속도 개선 방법' 정보과학회논문지: 시스템 및 이론, 제38권, 제6호, 페이지 275-280 (2011) 요약 및 페이지 277-279 참조.	1-17
A	여운구 외, 'BARM: 유전자 조립과 레퍼런스 얼라인먼트를 접목한 메타유전체 비닝 방법' 정보과학회논문지: 데이터베이스, 제38권, 제2호, 페이지 72-83 (2011) 전체 문헌 참조.	1-17
A	KR 10-2011-0049005 A (삼성에스디에스 주식회사) 2011.05.12 전체 문헌 참조.	1-17
A	US 2011-0270533 A1 (ZHANG 외) 2011.11.03 전체 문헌 참조.	1-17
A	박민서 외, '차세대 시퀀싱을 위한 짧은 서열 정렬에 대한 기술조사' 정보과학회논문지: 소프트웨어 및 응용, 제38권, 제10호, 페이지 503-511 (2011) 전체 문헌 참조.	1-17
☐ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: "A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 "E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 "L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 "O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 "P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 "T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 "X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. "Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. "&" 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2013년 10월 29일 (29.10.2013)		국제조사보고서 발송일 2013년 11월 01일 (01.11.2013)
ISA/KR의 명칭 및 우편주소  대한민국 특허청 (302-701) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-472-7140		심사관 김승범 전화번호 +82-42-481-3371 

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2011-0049005 A	2011/05/12	KR 10-1201626 B1	2012/11/08
US 2011-0270533 A1	2011/11/03	WO 2011-137368 A2	2011/11/03
		WO 2011-137368 A3	2012/02/16