(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0138466
(43) 공개일자 2019년12월13일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6869 (2018.01)
(52) CPC특허분류
C12Q 1/6869 (2018.05)
C12Q 2535/122 (2019.08)
(21) 출원번호 10-2018-0064950
(22) 출원일자 2018년06월05일
심사청구일자 2018년06월05일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
주식회사 마이크로바이오틱스
서울특별시 서초구 서초대로 397 ,A동5층509호 (서초동,부띠크모나코)
(72) 발명자
용동은
서울특별시 서대문구 연세로 50-1 연세의대진단검사의학교실
황연지
서울특별시 서대문구 연세로 50-1 연세의대 세균내성연구소
(뒷면에 계속)
(74) 대리인
이재영

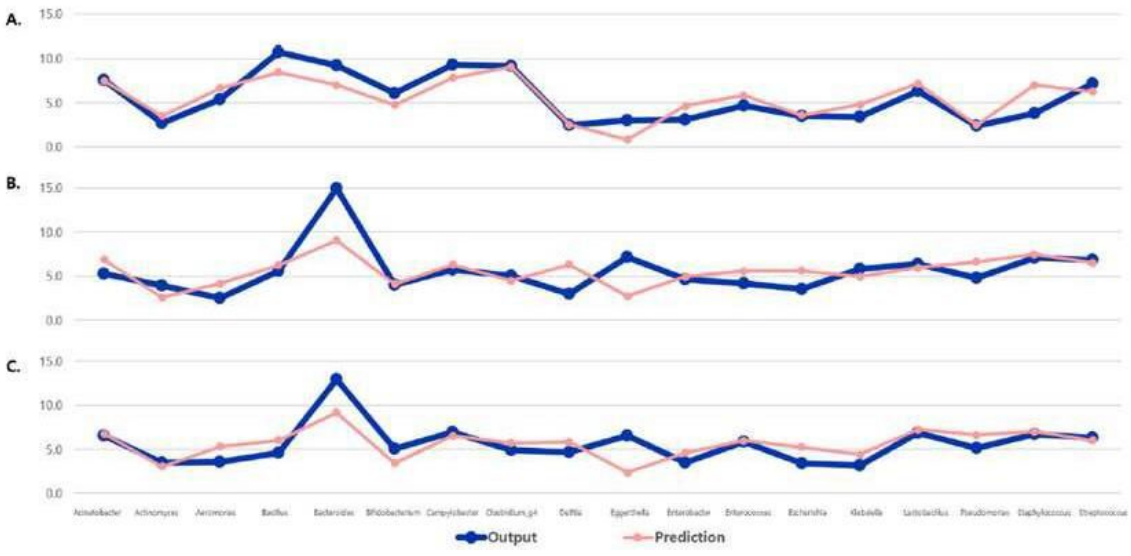
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **차세대 염기서열 분석법의 정확도를 분석하는 방법**

(57) 요 약

본 발명은 차세대 염기서열 분석법의 정확도를 분석하는 방법에 관한 것으로, 보다 상세하게는 서로 상이한 2종 이상의 박테리아로부터 추출된 게놈 DNA(genomic DNA)를 포함하는 게놈 DNA 인공 유전체를 준비하는 단계; 및 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석을 수행하는 단계;를 포함하며,상기 차세대 염기서열 분석 시 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 측정할 수 있다.

대 표 도



(72) 발명자

김주영

서울특별시 서대문구 연세로 50-1 연세의대 환경의생물학교실

문혜수

서울특별시 동대문구 한천로24길 74-8(장안동)

이 발명을 지원한 국가연구개발사업

과제고유번호 HI14C1324

부처명 보건복지부

연구관리전문기관 보건의료기술연구개발사업

연구사업명 글로벌 의료수요 해결을 위한 전략적 기술통합의 개방형 연구 비즈니스 플랫폼 구축

연구과제명 연세대학교 세브란스병원

기 여 율 1/1

주관기관 한국보건산업진흥원

연구기간 2014.10.01 ~ 2017.01.31

명 세 서

청구범위

청구항 1

서로 상이한 2종 이상의 박테리아로부터 추출된 게놈 DNA(genomic DNA)를 포함하는 게놈 DNA 인공 유전체를 준비하는 단계; 및

상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석을 수행하는 단계;를 포함하며,

상기 차세대 염기서열 분석 시 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 측정하는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 2

제1항에 있어서,

상기 박테리아는 아시네토박터(Acinetobacter) 속 박테리아, 악티노마이세스(Actinomyces) 속 박테리아, 아에로모나스(Aeromonas) 속 박테리아, 바실러스(Bacillus) 속 박테리아, 박테로이데스(Bacteroides) 속 박테리아, 비피도박테리움(Bifidobacterium) 속 박테리아, 캠필로박터(Campylobacter) 속 박테리아, 클로스트리듐(Clostridium) 속 박테리아, 델프티아(Delftia) 속 박테리아, 에게르텔라(Eggerthella) 속 박테리아, 엔테로박터(Enterobacter) 속 박테리아, 엔테로코커스(Enterococcus) 속 박테리아, 에스케리키아(Escherichia) 속 박테리아, 클렙시엘라 (Klebsiellla) 속 박테리아, 락토바실러스(Lactobacillus) 속 박테리아, 슈도모나스(Pseudomonas) 속 박테리아, 스타필로코커스(Staphylococcus) 속 박테리아 및 스트렙토코커스(Streptococcus) 속 박테리아로 이루어진 군에서 선택된 2종 이상을 포함하는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 3

제1항에 있어서,

상기 차세대 염기서열 분석 시 프라이머로 V1V2 영역에 대한 프라이머, V3V4 영역에 대한 프라이머 및 V6V8 영역에 대한 프라이머 중 2종 이상을 사용하여 수행한 뒤, 각 프라이머에 따라 분석된 각 박테리아 분포 비율을 비교하며 수행되는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 4

제1항에 있어서,

상기 상기 게놈 DNA 인공 유전체는 박테리아의 종류에 따라 분류된 제1 군 및 제2 군이 1:1~100의 농도 비율로 혼합된 것인, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 5

제4항에 있어서,

상기 차세대 염기서열 분석 시, 시료로 상기 제1 군 및 제2 군이 1:1의 농도 비율로 혼합된 것과, 상기 제1 군 및 제2 군이 1: 1 초과 100 이하의 농도 비율로 혼합된 것을 사용한 뒤, 각 시료에 따라 분석된 각 박테리아 분포 비율을 비교하며 수행되는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 6

제1항에 있어서,

상기 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 하기 식 1에 대입하여 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 예측 분포 비율(%)을 측정하는, 차세대 염기서열 분석법의 정확도를 분석하는 방법:

[식 1]

목표 박테리아의 예측 분포 비율(%) = A1 + A2 X (목표 박테리아의 시료의 양(㎕)) + A3 X (V3V4 영역의 GC 함량(%)) + A4 X (16S rRNA 유전자 복제 수(개)) + A5 X (게놈 사이즈(bp))

상기 식 1에서, A1은 16 내지 19이고, A2는 0.4 내지 0.7이며, A3는 -0.5 내지 -0.3이고, A4는 0.4 내지 0.6이며, A5는 -9E-07 내지 -4E-07이다.

청구항 7

제6항에 있어서,

상기 식 1에서 상기 A1은 16.40 내지 18.60이고, A2는 0.47 내지 0.56이며, A3는 -0.45 내지 -0.34이고, A4는 0.41 내지 0.52이며, A5는 -8.30E-07 내지 -4.81E-07인, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 8

제1항에 있어서,

상기 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수), 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%) 및 그람 양성 여부를 측정한 뒤 이들을 하기 식 2에 대입하여 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 예측 분포 비율(%)을 측정, 차세대 염기서열 분석법의 정확도를 분석하는 방법:

[식 2]

목표 박테리아의 예측 분포 비율(%) = A1 + A2 X (목표 박테리아의 시료의 양(㎕)) + A3 X (V3V4 영역의 GC 함량(%)) + A4 X (16S rRNA 유전자 복제 수(개수)) + A5 X (게놈 사이즈(bp)) + A6 X (그람 양성 여부)

상기 식 2에서, A1은 16 내지 19이고, A2는 0.4 내지 0.7이며, A3는 -0.5 내지 -0.3이고, A4는 0.4 내지 0.6이며, A5는 -9E-07 내지 -4E-07이고, A6는 -0.8 내지 -0.3이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0이다.

청구항 9

제8항에 있어서,

상기 식 2에서, A1은 16.40 내지 18.60이고, A2는 0.47 내지 0.56이며, A3는 -0.45 내지 -0.34이고, A4는 0.41 내지 0.52이며, A5는 -8.30E-07 내지 -4.81E-07이고, A6는 -0.74 내지 -0.34, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 10

제6항 또는 제8항에 있어서,

상기 목표 박테리아의 예측 분포 비율(%)을 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)

과 비교하여 차세대 염기서열 분석법의 정확도를 분석하는 단계를 더 포함하는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

청구항 11

제6항 또는 제8항에 있어서,

상기 차세대 염기서열 분석법 수행 시 농도가 상이한 2종류 이상의 시료 또는 2종류 이상의 프라이머를 이용하여 수행하고,

각 시료의 농도 또는 각 프라이머에 따른 목표 박테리아의 예측 분포 비율(%)을 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)과 비교하는 단계를 더 포함하는, 차세대 염기서열 분석법의 정확도를 분석하는 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 차세대 염기서열 분석법의 정확도를 분석하는 방법에 관한 것이다.

배 경 기 술

[0003] 지금까지 고효율 서열분석 기술(high-throughput sequencing technologies)이라고 알려진 차세대 염기서열 분석법(Next Generation Sequencing, NGS)이 급속도로 발전하였다. 미생물 실험에 대한 시각이 크게 변하였고, 특히 그 분석 비용이 매우 감소함에 따라 상기 차세대 염기서열 분석법의 응용이 확산되고 있다. 마이크로바이옴(microbiome)의 분석 시 미생물 군집의 특징화를 위하여 계통 발생적 연구를 위해 초가변 영역(hyper variable region)에 해당하는 16S rRNA 유전자의 서열 분석이 주로 수행되고 있다.

[0004] 뿐만 아니라 최근에는 마이크로바이옴의 분석을 위한 더 나은 플랫폼(platforms)이 제안되고는 있지만, 심층 시퀀싱(deep sequencing)으로부터 얻어지는 방대한 양의 데이터들이 실제 시료의 정보를 나타내고 있는 지에 대하여 여전히 의문이 제기되고 있다.

[0005] 하지만, 많은 연구에서 마이크로바이옴의 분석 시 NGS가 시퀀싱 데이터의 분석을 위한 표준화된 방법으로 활용할 수 없음에도 이를 고려하지 않고 진행되고 있다. 그 경우 시퀀싱 분석하는 전체 과정에서 많은 오류가 발생하고, 사용되는 NGS 플랫폼이나 소프트웨어 또는 데이터 베이스에 따라 서로 다른 결과를 나타내기도 한다. 상기한 마이크로바이옴의 분석을 특히 의료적 시료에 대하여 수행하는 경우 이러한 오류로 인하여 정확한 정보를 제공하지 못함에 따라 그 문제점이 더욱 중요시 되고 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명의 일 목적은 차세대 염기서열 분석법(Next Generation Sequencing, NGS)의 정확도를 분석하는 방법을 제공하고자 한다.

[0008] 본 발명의 다른 목적 및 이점은 하기의 발명의 상세한 설명, 청구범위 및 도면에 의해 보다 명확하게 된다.

과제의 해결 수단

[0010] 본 발명의 일 다른 구현 예에 따르면, 서로 상이한 2종 이상의 박테리아로부터 추출된 게놈 DNA(genomic DNA)를 포함하는 게놈 DNA 인공 유전체를 준비하는 단계; 및

[0011] 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석을 수행하는 단계를 포함하는, 차세대 염기서열 분석

법의 정확도를 분석하는 방법에 관한 것이다.

[0012] 본 발명에서 상기 게놈 DNA 인공 유전체는 서로 상이한 2종 이상의 박테리아로부터 추출된 게놈 DNA를 포함하면 되는 것이고, 상기 박테리아의 구체적인 종류를 제한하지 않으나, 예를 들면, 아시네토박터(Acinetobacter) 속 박테리아, 악티노마이세스(Actinomyces) 속 박테리아, 아에로모나스(Aeromonas) 속 박테리아, 바실러스(Bacillus) 속 박테리아, 박테로이데스(Bacteroides) 속 박테리아, 비피도박테리움(Bifidobacterium) 속 박테리아, 캠필로박터(Campylobacter) 속 박테리아, 클로스트리듐(Clostridium) 속 박테리아, 델프티아(Delftia) 속 박테리아, 에게르텔라(Eggerthella) 속 박테리아, 엔테로박터(Enterobacter) 속 박테리아, 엔테로코커스(Enterococcus) 속 박테리아, 에스케리키아(Escherichia) 속 박테리아, 클렙시엘라 (Klebsiellla) 속 박테리아, 락토바실러스(Lactobacillus) 속 박테리아, 슈도모나스(Pseudomonas) 속 박테리아, 스타필로코커스(Staphylococcus) 속 박테리아 및 스트렙토코커스(Streptococcus) 속 박테리아로 이루어진 군에서 선택된 2종 이상일 수 있다.

[0013] 또한, 본 발명에서 상기 박테리아는 아시네토박터 바우마니(Acinetobacter baumannii), 악티노마이세스 오돈톨리티쿠스(Actinomyces odontolylticus), 아에로모나스 하이드로필라(Aeromonas hydrophila), 바실러스 세레우스(Bacillus cereus), 박테로이데스 프라길리스(Bacteroides fragilis), 비피도박테리움 아돌레센티스(Bifidobacterium adolescentis), 캠필로박터 제주니(Campylobacter jejuni), 클로스트리듐 디피실리(Clostridium difficile), 델프티아 애시도보란스(Delftia acidovorans), 에게르텔라 렌타(Eggerthella lenta), 엔테로박터 클로아케(Enterobacter cloacae), 엔테로코커스 페컬리스(Enterococcus faecalis), 에스케리키아 콜라이(Escherichia coli), 클렙시엘라 뉴모니아(Klebsiellla pneumonia), 락토바실러스 퍼멘텀(Lactobacillus fermentum), 슈도모나스 에루지노사(Pseudomonas aeruginosa), 스타필로코커스 아우레우스(Staphylococcus aureus) 및 스트렙토코커스 뉴모니아(Streptococcus pneumonia)로 이루어진 군에서 선택된 2종 이상일 수 있다. 이때 상기 각 박테리아는 앞서 열거된 각 속의 박테리아의 어느 대표적 일 예시를 나타낸 것이고, 이에 제한되는 것은 아니다.

[0014] 본 발명의 게놈 DNA 인공 유전체에서, 상기 박테리아는 상기 열거된 아시네토박터 바우마니(Acinetobacter baumannii), 악티노마이세스 오돈톨리티쿠스(Actinomyces odontolylticus), 아에로모나스 하이드로필라(Aeromonas hydrophila), 바실러스 세레우스(Bacillus cereus), 박테로이데스 프라길리스(Bacteroides fragilis), 비피도박테리움 아돌레센티스(Bifidobacterium adolescentis), 캠필로박터 제주니(Campylobacter jejuni), 클로스트리듐 디피실리(Clostridium difficile), 델프티아 애시도보란스(Delftia acidovorans), 에게르텔라 렌타(Eggerthella lenta), 엔테로박터 클로아케(Enterobacter cloacae), 엔테로코커스 페컬리스(Enterococcus faecalis), 에스케리키아 콜라이(Escherichia coli), 클렙시엘라 뉴모니아(Klebsiellla pneumonia), 락토바실러스 퍼멘텀(Lactobacillus fermentum), 슈도모나스 에루지노사(Pseudomonas aeruginosa), 스타필로코커스 아우레우스(Staphylococcus aureus) 및 스트렙토코커스 뉴모니아(Streptococcus pneumonia) 외에도 추가의 박테리아를 더 포함할 수 있는 것으로, 제한되지 않는다.

[0015] 본 발명에서 상기 게놈 DNA 인공 유전체는 차세대 염기서열 분석 시 목표 박테리아의 시료 투입량(㎕); 각 박테리아의 게놈 사이즈(bp), 16S rRNA 유전자의 복제 수(개수) 또는 상기 16S rRNA 유전자 내 GC 함량(%) 등과 같은 박테리아 특성; 및 프라이머;가 미치는 영향을 확인하기 위한 것이다.

[0016] 본 발명에서 상기 박테리아로부터 게놈 DNA 추출 키트로, 예를 들어, GenEluteTM Bacterial Genomic DNA kit (Sigma, USA)를 사용하여 추출된 게놈 DNA를 포함할 수 있으나, 이에 제한되는 것은 아니다. 이때 상기와 같이 박테리아로부터 게놈 DNA 추출 시 그람-양성(Gram-positive) 박테리아에 대한 추출 방법에 의할 수 있다.

[0017] 또한, 본 발명에서 상기 게놈 DNA 인공 유전체는, 기타 오염을 방지하기 위하여 추출된 게놈 DNA를 정제한 것이 바람직하다.

[0018] 본 발명에서 상기 차세대 염기서열 분석 시 시료로 사용되는 게놈 DNA 인공 유전체를 박테리아의 종류에 따라 임의의 두 군으로 분류한 뒤 이들을 1:1~100의 농도 비율로 혼합한 것을 사용함으로써 투입하는 시료의 농도가 차세대 염기서열 분석법에 미치는 영향을 분석할 수 있다. 예를 들면, 상기 열거한 18종의 박테리아로부터 추출된 게놈 DNA를 포함하는 게놈 DNA 인공 유전체에 있어서, 박테리아의 종류에 따라 임의의 두 군으로 분류할 수 있고, 바람직하게는 18종의 박테리아 중 1~9종의 박테리아를 포함하는 제1 군과, 나머지 박테리아를 포함하는 제2 군으로 분류한 뒤, 이들을 1:1~100의 농도 비율로 혼합한 것을 시료로 사용할 수 있다.

[0019] 본 발명에서 상기 게놈 DNA 인공 유전체로 제1 군과 제 2군이 1:1~100의 농도 비율로 혼합한 것을 사용하는 경

우, 바람직하게는 상기 제1 군과 제2 군의 1:1 혼합물(농도비); 및 1:2 혼합물, 1:4 혼합물, 1:10 혼합물 및 1:100 혼합물 중 1종 이상;을 시료로 사용한 뒤, 각 시료에 따른 분석 결과물(각 박테리아의 분포 비율)을 비교함으로써 시료 농도가 상기 차세대 염기서열 분석법에 미치는 영향을 분석할 수 있다.

[0020] 본 발명에서 상기 차세대 염기서열 분석 후 얻어진 결과물을 Mothur- Silva 데이터베이스, Mothur-Eztaxon 데이터 베이스 및 BaseSpace-Greengenes 중 1종 이상을 사용하여 분석한 뒤 상기 조성물 내 각 박테리아의 실제 분포 비율과 비교함으로써 차세대 염기서열 분석 후 얻어진 결과물을 해석하는 데이터베이스가 차세대 염기서열 분석에 미치는 영향을 분석할 수 있다.

[0021] 또한 본 발명에서 상기 차세대 염기서열 분석 후 얻어진 결과물을 Mothur-Silva 데이터베이스, Mothur-Eztaxon 데이터 베이스, 및 BaseSpace-Greengenes 중 2종 이상을 사용하여 분석한 뒤 각 데이터 베이스에 따른 각 박테리아의 분포 비율을 서로 비교함으로써 각 데이터 베이스에 따른 각 박테리아의 분포 비율을 비교함으로써 DNA 추출 방법이 상기 차세대 염기서열 분석에 미치는 영향을 분석할 수 있다.

[0022] 본 발명에서는 상기 차세대 염기서열 분석에 의하여, 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 측정할 수 있다.

[0023] 본 발명에서는 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 하기 식 1에 대입하여 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 예측 분포 비율(%)을 측정할 수 있다:

[0025] [식 1]

[0026] 목표 박테리아의 예측 분포 비율(%) = A1 + A2 X (목표 박테리아의 시료의 양(㎕)) + A3 X (V3V4 영역의 GC 함량(%)) + A4 X (16S rRNA 유전자 복제 수(개)) + A5 X (게놈 사이즈(bp))

[0027] 상기 식 1에서, A1은 16 내지 19이고, A2는 0.4 내지 0.7이며, A3는 -0.5 내지 -0.3이고, A4는 0.4 내지 0.6이며, A5는 -9E-07 내지 -4E-07이다.

[0028] 본 발명에서 상기 식 1에서, A1은 16.40 내지 18.60이고, A2는 0.47 내지 0.56이며, A3는 -0.45 내지 -0.34이고, A4는 0.41 내지 0.52이며, A5는 -8.30E-07 내지 -4.81E-07일 수 있다.

[0029] 본 발명에서 상기 식 1에서, A1은 16.43 내지 18.557이고, A2는 0.471 내지 0.555이며, A3는 -0.431 내지 -0.343이고, A4는 0.415 내지 0.515이며, A5는 -8.292E-07 내지 -4.816E-07일 수 있다.

[0030] 본 발명에서 상기 식 1에서 A1은 17.327이고, A2는 0.53375이며, A3는 -0.388이고, A4는 0.46125이며, A5는 -6.17075E-07일 수 있다.

[0031] 또한, 본 발명에서 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수), 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%) 및 그람 양성 여부를 측정한 뒤 이들을 하기 식 2에 대입하여 상기 게놈 DNA 인공 유전체 내 목표 박테리아의 예측 분포 비율(%)을 측정할 수 있다:

[0033] [식 2]

[0034] 목표 박테리아의 예측 분포 비율(%) = A1 + A2 X (목표 박테리아의 시료의 양(㎕)) + A3 X (V3V4 영역의 GC 함량(%)) + A4 X (16S rRNA 유전자 복제 수(개수)) + A5 X (게놈 사이즈(bp)) + A6 X (그람 양성 여부)

[0035] 상기 식 2에서, A1은 16 내지 19이고, A2는 0.4 내지 0.7이며, A3는 -0.5 내지 -0.3이고, A4는 0.4 내지 0.6이며, A5는 -9E-07 내지 -4E-07이고, A6는 -0.8 내지 -0.3이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0이다.

[0036] 본 발명에서 상기 식 2에서, A1은 16.40 내지 18.60이고, A2는 0.47 내지 0.56이며, A3는 -0.45 내지 -0.34이고, A4는 0.41 내지 0.52이며, A5는 -8.30E-07 내지 -4.81E-07이고, A6는 -0.74 내지 -0.34이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0일 수 있다.

[0037] 본 발명에서 상기 식 2에서, A1은 16.43 내지 18.557이고, A2는 0.471 내지 0.555이며, A3는 -0.431 내지

-0.343이고, A4는 0.451 내지 0.515이며, A5는 -8.292E-07 내지 -4.816E-07이고, A6는 -0.731 내지 -0.34이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0일 수 있다.

[0038] 본 발명에서 상기 식 2에서 A1은 17.327이고, A2는 0.53375이며, A3는 -0.388이고, A4는 0.46125이며, A5는 -6.17075E-07이고, A6는 -0.49775일 수 있다.

[0040] 본 발명에서는 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석법을 수행한 뒤 측정된 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 상기 식 1 또는 2에 대입하여 얻어진 목표 박테리아의 예측 분포 비율(%)을 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)과 비교하는 단계를 수행하여 차세대 염기서열 분석법의 정확도를 분석할 수 있다.

[0041] 또한, 본 발명에서는 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석을 수행할 때 농도가 상이한 2종류 이상의 시료를 사용하여 수행한 뒤 측정된 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 상기 식 1 또는 2에 대입하여 얻어진 각 시료에 따른 목표 박테리아의 예측 분포 비율(%)을, 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)과 비교하는 단계를 수행하여 각 시료의 농도가 차세대 염기서열 분석법의 정확도에 미치는 영향을 분석할 수 있다.

[0042] 또한, 본 발명에서는 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석법을 수행할 때 서로 상이한 2종류 이상의 프라이머를 사용하여 수행한 뒤 측정된 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 상기 식 1 또는 2에 대입하여 얻어진, 각 프라이머에 따른 목표 박테리아의 예측 분포 비율(%)을 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)과 비교하는 단계를 수행하여 각 DNA 추출 방법이 차세대 염기서열 분석법의 정확도에 미치는 영향을 분석할 수 있다.

[0043] 또한, 본 발명에서는 상기 게놈 DNA 인공 유전체에 대하여 차세대 염기서열 분석을 수행할 때 2종류 이상의 차세대 염기서열 분석 데이터 베이스를 사용하여 수행한 뒤, 측정된 목표 박테리아의 시료의 양(㎕), 게놈 사이즈(bp), 16S rRNA 유전자 복제 수(개수) 및 상기 16S rRNA 유전자 내 V3V4 영역의 GC 함량(%)을 상기 식 1 또는 2에 대입하여 얻어진 각 데이터 베이스에 따른 목표 박테리아의 예측 분포 비율(%)을 상기 목표 박테리아의 게놈 DNA 인공 유전체 내 실제 분포 비율(%)과 비교하는 단계를 수행하여 각 데이터 베이스가 차세대 염기서열 분석법의 정확도에 미치는 영향을 분석할 수 있다.

발명의 효과

[0045] 본 발명에서는 차세대 염기서열 분석법의 정확도에 오류를 일으킬 수 있는 인자를 분석하는 방법에 관한 것이다.

도면의 간단한 설명

[0047] 도 1은 본 발명의 일 실시예에 따른 인공 유전체 조성물에 포함되는 박테리아 균주를 나타낸 것이다.

도 2는 본 발명의 일 실시예에 따른 인공 유전체 조성물에 포함되는 각 박테리아 균주에 V1V2 프라이머, V3V4 프라이머, V6V8 프라이머의 결합을 나타낸 것이다.

도 3은 본 발명의 일 실시예에 있어서, V1V2 프라이머, V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭된 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 있어서 박테리아의 빈도를 문(phylum) 수준으로 분석한 결과를 나타낸 것이다.

도 4는 본 발명의 일 실시예에 있어서, V1V2 프라이머, V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭된 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 있어서 박테리아의 빈도를 속(genus) 수준으로 분석한 결과를 나타낸 것이다.

도 5는 본 발명의 일 실시예에 있어서, V1V2 프라이머, V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭된 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체의 히트맵(heatmap)과 계통도(dendrogram)를 나타낸 것으로, 붉은색은 게놈 DNA 인공 유전체, 녹색은 플라스미드 인공 유전체, 청색은 PCR 인공 유전체를 나타낸다.

도 6은 본 발명의 일 실시예에 있어서, V1V2 프라이머, V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭된 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 있어서 예측되는 비율(Expected)과 비교한 거리 매트릭스의 PCA 결과를 나타낸 것이다. PCA를 위하여 Bray-Curtis 비유사성 인덱스(dissimilarity index)가 사용되었다. 모양은 프라이머의 종류에 따라 구분되고, 색은 인공 유전체의 종류에 따라 구분된다.

도 7은 본 발명의 일 실시예에 있어서, 세포 인공 유전체에 있어서 4종류의 DNA 추출 키트에 따라 박테리아의 빈도를 문(phylum) 수준으로 분석한 결과를 나타낸 것이다.

도 8은 본 발명의 일 실시예에 있어서, 세포 인공 유전체에 있어서 4종류의 DNA 추출 키트에 따라 박테리아의 빈도를 문(phylum) 수준으로 분석한 결과를 나타낸 것이다.

도 9는 본 발명의 일 실시예에 있어서, 4종류의 DNA 추출 키트를 사용하여 추출된 DNA를 포함하는 세포 인공 유전체를 V1V2 프라이머, V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭시킨 뒤, DNA 추출 키트 및 프라이머의 종류에 따라 분석한 히트맵(heatmap)과 계통도(dendrogram)를 나타낸 것이다.

도 10은 본 발명의 일 실시예에 있어서, 4종류의 DNA 추출 키트를 사용하여 추출된 DNA를 V3V4 프라이머 또는 V6V8 프라이머를 사용하여 증폭시킨 세포 인공 유전체에 있어서, 예측되는 비율(Expected)과 비교한 거리 매트릭스의 PCA 결과를 나타낸 것이다. PCA를 위하여 Bray-Curtis 비유사성 인덱스(dissimilarity index)가 사용되었다. 모양은 프라이머의 종류에 따라 구분되고, 색은 DNA 추출 키트의 종류에 따라 구분된다.

도 11은 본 발명의 일 실시예에 있어서, 게놈 DNA 인공 유전체의 시퀀싱 결과물에 박테리아 특성이 미치는 영향을 그래프로 나타낸 것이다.

도 12의 (A), (B) 및 (C) 각각은 본 발명의 일 실시예에 있어서, 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 대하여 각 박테리아의 실제 빈도(Output)와 예측되는 빈도(Prediction)를 비교한 그래프를 나타낸 것이다.

도 13은 본 발명의 일 실시예에 있어서, 차세대 염기서열 분석 시 시료 투입량에서 정량적 변화의 영향을 나타낸 것이다.

도 14는 본 발명의 일 실시예에 있어서, 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에서 3종류의 상이한 파이프라인 및 데이터베이스로 분석하여 얻어진 박테리아 빈도를 비교한 그래프를 나타낸 것이다.

도 15는 본 발명의 일 실시예에 있어서, EzTaxon 데이터베이스를 이용하여 박테리아를 종 수준으로 동정한 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0048] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로서, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[0050] **실시예**

[0052] 1. 인공 유전체 미생물의 배양

[0053] 폐와 소화관(gut)에서 공생하는 박테리아로 알려진 도 1의 18종의 박테리아 균주를 준비하였다. 18종의 박테리아 균주를 성장 조건을 고려하여 배지 상에서 배양하였다.

[0054] 하기 균주 중 조건적 혐기성 균주(Facultative anaerobic strains)의 경우 혈액 한천 배지(BAP; Asan Pharmaceutical, Korea)에서 37℃의 온도 조건 하에서 24시간 동안 배양하였다. 절대 혐기성 균주(Obligate

anaerobes)의 경우 브루셀라 한천 플레이트(Brucella agar plates)(Asan Pharmaceutical, Korea)에서 37℃의 온도, 및 질소 가스, 이산화탄소 가스 및 산소의 분위기 하에서 48시간 동안 배양하였다. 단, 캠필로박터 제주니의 경우 BAP 배지에서 40℃의 온도 및 미호기성 조건(microaerophilic condition) 하에서 48시간동안 배양하였고, 락토바실러스 퍼멘텀의 경우 조건적 혐기성 균주에 속하지만, MRS 아가 (BD)에서 5% CO_2 인큐베이터 및 37℃의 온도 하에서 24시간 동안 배양하였다.

[0056] 2. 박테리아 종 동정

[0057] 오염에 주의하기 위하여, 사용에 앞서 모든 박테리아 배양물을 동정하였다. 구체적으로는 MALDI-TOF MS(matrix-assisted laser desorption/ionization time-of-flight mass spectrometer)(Bruker Daltonics, France)로 박테리아 종을 확인하기 위하여, 단일 박테리아 콜로니를 멸균된 루프(autoclaved loop)로 취한 뒤 상기 MALDI-TOF MS 플레이트의 한 부분에 도말한 후 70% 포름산 1 μl를 첨가하였다. 포름산이 건조한 후 매트릭스 (alpha-cyano-4-hydroxycinnamic acid [HCCA]; Sigma, USA) 1 μl를 동일 부위에 첨가한 후 완전히 건조한 뒤에 플레이트를 장치에 삽입하였다. 다른 동정 방법으로는 PCR을 이용하여 16S rRNA 유전자를 증폭하였다. 보다 상세하게는 Takara Taq DNA 중합 효소 0.25 μl, 10X PCR 버퍼 5 μl, 데옥시리보핵산 (Dntp) 혼합물 4 μl, 각각의 프라이머(10 μM) 2 μl, 박테리아 DNA 및 PCR-grade water 1μl로 구성된 50 μl의 Takara Taq 키트(Takara Bio Inc., Japan)를 사용하였다. PCR은 95℃에서 5분, 95℃에서 1분, 55℃에서 30초, 72℃에서 40초의 30 사이클; 및 72℃에서 10분 동안 수행하였다. 이후 얻어진 PCR 산물을 PCR 정제 키트(Qiagen, Germany)로 정제한 뒤 35 μl를 분리하였다. 정제된 PCR 산물을 생거(Sanger)법으로 염기서열을 분석한 뒤 EzBioCloud 웹사이트(ChunLab, Korea) 상의 Eztaxon 데이터 베이스를 사용하여 박테리아를 동정하였다.

[0059] 3. 세포 인공 유전체의 제작

[0060] 상기 1. 에서 준비한 박테리아 세포를 한천 플레이트에서 배양한 뒤, 캠필로박터 제주니의 경우 LB 브로쓰 배지에서 40℃의 온도 하에서 계대 배양하였고, 그 외의 박테리아의 경우 각 박테리아 성장 조건에 따라 37℃의 온도 하에서 16~24 시간 동안 계대 배양하였다. 각 박테리아 세포의 양이 고르게 분포될 수 있도록 OD600를 측정하여 그 값이 0.03이 되도록 희석한 뒤 neubauer 챔버(Marienfeld Superior, Germany)에서 광학 현미경을 이용하여 박테리아 세포수를 계수하였다. 각 박테리아 세포의 수는 1 X 10^7 내지 5 X 10^7 cells/ml의 범위 하에서 측정되었다. 각 박테리아 세포 1ml씩 하나의 튜브에 첨가하여 인공 유전체를 제조하였고, DNA 추출을 위하여 총 12개의 세포 인공 유전체를 제조하였다. 세포 인공 유전체의 총 3배수를 각 DNA 추출 방법에 사용하므로, 이하의 실험에서는 총 4개의 DNA 추출 키트를 사용하였다. 상기 DNA 추출 키트로는 하기 표 1에 나타낸 바와 같이, 상업적으로 판매되고 있는 MP Bio Fast Soil kit (MP), Qiagen Stool Mini Kit (QiaS), Qiagen Blood and Tissue kit (QiaB) 및 Sigma GenEluteTM Bacterial Genomic DNA kit (Sig)를 사용하였다. 용출(elution)을 위하여 각 키트 내 용출 버퍼 100μl를 사용하였고, PCR grade water는 음성 대조군(blank control)으로 사용하였다. Nanodrop 및 Quantus를 이용하여 추출된 DNA의 질과 양을 평가하였다.

표 1

[0062]

		적용(Application)	용균(Lysis)		용출(Elution)	
MP	Fast Soil Kit	땅(Soil)	화학적 방법	Bead-beating	컬럼(Column)	50-100㎕
QiaB	Blood & Tissue kit	혈액, 조직	화학적 방법	95℃	컬럼	100-200㎕
QiaS	Stool Mini kit	대변(Stool)	화학적 방법	95℃	컬럼	200㎕
Sig	GenElute bacterial DNA kit	박테리아	화학적 방법	효소(리소자임)	컬럼	100-200㎕

[0064] 4. 게놈 DNA, 플라스미드 및 PCR 인공 유전체의 제작

[0065] 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체를 제작하기 위하여, GenEluteTM 박테리아 게놈 DNA 키트 (Sigma, USA)를 이용하여 총 18종의 박테리아 균주로부터 게놈 DNA(gDNA)를 추출하고, 그람-양성 박테리아 준비 공정을 수행하였다. 추출한 gDNA를 Genomic DNA Clean & ConcentratorTM-25 (Zymo Research, USA)와 DNA 추출 키트에 포함되어 있던 RNase 20 μl를 사용하여 정제하였다. 정제된 gDNA를 1 kb 래더(ladder)와 함께 1% 아가로스 겔에 로딩한 뒤 DNA 분해 상태를 확인하였다. 이후 Nanodrop (Life Technologies, USA), 형광 광도계(fluorometer) 및 Quantus (Promega, USA)를 이용하여 양 및 질을 확인하였다. RNA 오염 여부는 18종의 박테리아 gDNA를 모세관 전기이동(capillary electrophoresis)하여 확인하였다.

[0066] 플라스미드 인공 유전체 및 PCR 인공 유전체의 제작을 위하여 보편적 프라이머인 27F 및 1492R을 이용하여 16S rRNA 유전자를 얻었다. 단, 시약 혼합물로는 Takara Taq DNA 중합 효소 0.25 μl, 10X PCR 버퍼 5 μl, 데옥시리보핵산 (Dntp) 혼합물 4 μl, 각각의 프라이머(10 μM) 2 μl, 박테리아 DNA 및 PCR-grade water 1μl로 구성된 50 μl의 Takara Taq 키트(Takara Bio Inc., Japan)를 사용하여, 95℃에서 5분, 95℃에서 1분, 55℃에서 30초, 72℃에서 40초의 30 사이클; 및 72℃에서 10분 동안 PCR을 수행하고, 4℃에서 유지하였다. PCR 증폭 산물은 PCR 정제 키트(Qiagen, Germany)를 사용하여 정제한 뒤 EB 버퍼 35 μl를 이용하여 용출시켰다. 정제된 PCR 산물을 PCR 인공 유전체로 사용하였다. PCR 산물을 1% 아가로스 겔에 로딩한 뒤 1,500bp에서 단일 밴드를 확인하였다.

[0067] TOPcloner PCR 클로닝 키트 (Enzynomics, Korea)를 이용하여 상기와 같이 정제된 16S rRNA 유전자 앰플리콘 산물을 삽입물로 하고, DH5-alpha를 형질 전환을 위한 컴피턴트 세포(competent cell)로 하여 클로닝(cloning)을 수행하였다. 16S rRNA 유전자가 벡터에 잘 삽입되었는지 확인하기 위하여, 증폭을 위해 콜로니를 취한 뒤 플라스미드 추출을 위하여 카나마이신(kanamycin)(50 μg/ml)이 첨가된 LB 브로쓰에서 계대 배양하였다. QIAprep Miniprep kit (Qiagen, Germany)와 용출을 위한 EB 버퍼 35 μl를 이용하여 플라스미드를 정제한 뒤 플라스미드 인공 유전체로 사용하기 전까지 -20℃에서 보관하였다. 복제된 플라스미드는 품질을 위하여 1% 아가로스 겔에 담궈 놓았다.

[0068] 상기와 같이 준비한 총 3가지 인공 유전체로, 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 있어서, 각 박테리아 균주별 산물을 10 μl씩 취한 뒤 하나의 튜브에 첨가하여 20ng/μl의 농도가 되도록 희석하여 인공 유전체를 제작하였다. 이하 차세대 염기서열 분석법에서 투입량(input)의 영향을 확인하기 위하여 18종 박테리아의 인공 유전체를 임의로 2그룹, 즉 그룹 A와 그룹 B로 분류한 뒤 이들을 1:1, 1:2, 1:4, 1:10 및 1:100의 비율로 혼합하였다.

[0070] 5. 인 실리코(in silico) 프라이머 선별

[0071] Illumina MiSeq 시퀀싱 플랫폼에서 16S rRNA 유전자를 증폭하기 위한 PCR 프라이머를 Geneious R9.1 인 실리코로 확인하였다. 타겟 부위를 선별한 뒤, 하기 표 2에 나타낸 V1V2, V3V4 및 V6V8 부위에 대한 프라이머를 사용하였다.

표 2

[0073]

부위 및 프라이머	서열(5'-3')	서열번호
V1V2	정방향: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGRGTTYGATYMTGGCTCAG	서열번호 1
	역방향: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGCCTCCCGTAGGAGT	서열번호 2
V3V4	정방향: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG	서열번호 3
	역방향: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC	서열번호 4
V6V8	정방향: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAACTYAAAKRAATWGACGG	서열번호 5
	역방향: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGGGCGGTGWGTAC	서열번호 6

[0075] 단, 상기 표 2에서 서열번호 1 내지 6은 Universal ambiguity code에 근거한 것으로, 각 핵산 코드는 하기 표 3에 정의된 바와 같다.

표 3

[0077]

핵산 코드	염기	상보 핵산
A	아데닌(Adenine)	T
C	구아닌(Guanine)	C
G	시토신(Cytosine)	G
T	티민(Thymine)	A
Y	피리미딘(C 또는 T)	R
R	퓨린(A 또는 G)	Y
W	A 또는 T	W
S	G 또는 C	S
K	케토(T 또는 G)	M
M	아미노(C 또는 A)	K
D	A, G 또는 T(C는 아님)	H
V	A, C 또는 G(T는 아님)	B
H	A, C 또는 T(G는 아님)	D
B	C, G 또는 T(A는 아님)	V
X/N	모든 염기	X/N

[0079] 6. 차세대 염기서열 분석

[0080] 상기와 같이 준비된 인공 유전체를 라이브러리로 준비한 뒤, 앰플리콘(amplicon) PCR, PCR 산물 클리닝(cleaning), 인덱스(index) PCR, PCR 산물 클리닝, 정규화, 풀링(pooling), 변형(denaturation) 및 희석(dilution)을 포함하는 Illumina 16S Metagenomic Sequencing Library Preparation guide(Turnbaugh, P. J. Quince, C. Faith, J. J. McHardy, A. C. Yatsunenko, T. Niazi, F. et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci USA 2010; 107(16):7503-8)를 수행하였다. 즉, 2X KAPA HiFi HotStart ReadyMix (Roche, Switzerland) 12 μl, 각 오버행(overhang) 어댑터(adapter)가 첨가된 프라이머(1 μM) 5 μl, 샘플 DNA 3 μl로 구성된 반응 혼합물 25 μl를 이용하여 앰플리콘 PCR을 수행하였다. 이때, 16S rRNA 유전자 중 V1V2, V3V4 및 V6V8의 서로 상이한 3개의 부위를 타겟으로 하는 상기 표 2의 3종의 프라이머를 사용하였다. 인덱스 PCR을 위하여, 2X KAPA HiFi HotStart ReadyMix 25 μl, Nextera XT Index 키트 (Illumina) 유래 각 인덱스 프라이머 5 μl, 앰플리콘 PCR 산물 5 μl, 및 PCR grade water 10 μl로 구성된 반응 혼합물 50 μl를 사용하였다. 각 PCR 공정을 위하여, Agentcourt AMPure XP beads (Beckman Coulter, USA)를 이용하여 PCR 산물을 정제하였다. 증폭 및 정제된 시료를 4 nM까지 정규화한 뒤, 하나의 튜브에 풀링(pooling)하였다. Library 및 PhiX Control v3 키트(Illumina)를 변형시킨 뒤 각각 6 pM 및 12.5 pM으로 희석하고, 이들을 3:1의 비율로 혼합하였다. 라이브러리는 Illumina MiSeq 시퀀서(sequencer)에서 V3 600 cycle 키트(Illumina)를 이용하여 시퀀싱하였다.

[0081] Illumina MiSeq로부터 얻은 데이터를 페어드 엔드 리드(paired-end reads)가 조립되도록 하여 Mothur v1.39.5. Contigs로 진행하였고, EzTaxon 데이터베이스 (ChunLab, Korea)와 정렬하였다. UCHIME를 이용하여 키메라를 제거하고 남은 서열은 EzTaxon를 참고하여 분류하였다. 생물정보학(bioinformatics) 파이프라인을 비교하기 위하여 contigs의 정렬을 위해 Mothur에서 Silva v128를 사용하였고, 분류 시 참고를 위하여 RDP v9를 사용하였다. Greengenes v13에서 얻어진 분류 자료와 비교하기 위하여 Illumina BaseSpace, 16S 메타제노믹스 어플리케이션(metagenomics application)을 사용하였다.

[0083] 7. 결과

[0084] (1) 인공 유전체의 양 및 질 평가 결과

[0085] 박테리아 세포, 게놈 DNA, 16S rRNA 유전자 클로닝된 플라스미드, 16S rRNA 유전자 앰플리콘과 같은 박테리아

산물을 이용하여 세포 인공 유전체, 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체의 총 4가지 인공 유전체를 제작하였다. 각 박테리아 산물의 양과 질은 심층 분석 시 RNA 또는 다른 종의 오염이 존재하는 지 여부와 관련하여 중대한 역할을 하므로, 상기 4가지 mock 시료를 모두 평가하였다. 박테리아 세포를 제외하고 모든 박테리아 산물에 있어서 농도를 Quantus (Promega)로 측정하였고, 농도는 모두 30 ng/μl를 초과하였다. 세포 인공 유전체로부터 추출한 DNA 농도는 10 ng/μl를 초과하지 않았다.

[0086] 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체와 2개의 비-주형 컨트롤(NTC) 샘플을 3개의 프라이머를 사용하여 시퀀싱하였다. 박테리아 세포 인공 유전체의 경우 서로 상이한 4개의 DNA 추출 키트를 사용하여 3번 시퀀싱 하였다. 그 결과 하기 표 4에서 보는 바와 같이, NTC 샘플은 평균 329 리드를 생산하였으나, 인공 유전체는 최소 33062 내지 최대 139577의 리드를 생산하였다. Qiagen Blood & Tissue 키트 (QiaB)를 이용하여 박테리아 세포 인공 유전체로부터 추출한 DNA의 경우 타겟 유전자에 대하여 83 리드만을 생산할 수 있었는 바, 크게 증폭시키지 못하는 것을 알 수 있었다. 단, QiaB로 추출된 DNA의 경우 모든 인공 유전체에 있어서 신뢰할 수 있는 정도의 리드 카운트(read counts)를 형성함을 알 수 있었다.

표 4

[0088]

인공 유전체	타켓 영역에서 평균 리드 카운트 수		
	V1V2	V3V4	V6V8
gDNA mock[a]	66299	106807.4	41342
Plasmid mock[a]	66098	82320.8	42732
PCR mock[a]	59503.5	68360.4	30104
Cell mock[b]			
MP Bio Soil Kit	69793.7	107926.3	38284
Qiagen Stool Kit	44108.7	67297.7	83
Qiagen Blood & Tissue Kit	66129.3	95239	38623.7
Sigma Bacterial gDNA Kit	55277.7	129354.7	56697.7
a Samples were sequenced in duplicates. b Samples were sequenced in triplicates.			

[0090] (2) 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체를 이용한 프라이머의 영향 분석

[0091] 차세대 염기서열 분석법에서 사용되는 프라이머가 미치는 영향을 확인하기 위하여 16S rRNA 유전자의 서로 상이한 영역을 증폭시키는 3종의 프라이머를 사용하였다. 박테리아 DNA에 대한 프라이머 미스매치의 경우, 박테리아 16S rRNA 유전자에 대한 3종의 프라이머의 인 실리코 분석을 수행하였다. 3개의 프라이머 세트는 NCBI 뉴클레오타이드 데이터베이스에서 얻은 18종 박테리아의 16S rRNA 유전자 서열과 매칭된다(도 2).

[0092] 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체를 문 수준(phylum level)에서 분석하였다(도 3). 인공 유전체는 총 18종의 박테리아 균주로 구성되는데, 이들은 4가지 문에 포함되며, 프로테오박테리아(Proteobacteria) 8 균주, 피르미쿠테스(Firmicutes) 6 균주, 박테로이데스(Bacteroides) 1 균주, 악티노박테리아(Actinobacteria) 3 균주로 구성되어 있다. 3개의 프라이머로 증폭된 gDNA mock은 피르미쿠테스가 가장 높은 빈도(abundance)를 보였고, 그 다음은 프로테오박테리아에 해당하였다. 플라스미드 mock과 PCR mock은 V1V2 프라이머와 V3V4 프라이머를 사용한 경우에서는 피르미쿠테스가 프로테오박테리아 보다 더욱 높은 비율을 나타내었으나, V6V8 프라이머를 사용한 경우에는 프로테오박테리아가 피르미쿠테스보다 높은 비율을 보였다. 또한, 3개의 인공 유전체에 있어서, 박테로이데스 문이 V1V2 프라이머와 V3V4 프라이머를 사용했을 때 가장 적은 빈도를 보였고, V6V8 프라이머 세트를 사용하였을 때는 악티노박테리아가 가장 낮은 빈도를 보였다.

[0093] 도 4에서 보는 바와 같이, 속 수준(genus level)에 있어서는 V1V2 프라이머와 V3V4 프라이머를 사용했을 때 에게르텔라(Eggerthella)는 확인되지 않았으며, 슈도모나스는 그 빈도가 각각 0.03%, 0.01% 및 0.11%로 매우 낮은 수준으로 검출되었다. V1V2 프라이머를 사용하였을 때 가장 높은 빈도를 나타낸 종으로는 각각 클로스트리디움(Clostridium), 락토바실러스(Lactobacillus) 및 캠필로박터(Campylobacter)인데, V3V4 프라이머를 사용했을

때는 박테로이데스를 제외하고는 높은 빈도 순위가 일치하지 않았다. V6V8 프라이머를 사용하였을 때 가장 높은 빈도의 3개의 종은 클로스트리디움, 캠필로박터 및 박테로이데스에 속하였다. V3V4 영역을 타겟으로 했을 때 엔테로박터(Enterobacter), 델프티아(Delftia) 및 악티노마이세스(Actinomyces)는 낮은 빈도를 보였으나, V6V8 영역을 타겟으로 하였을 때는 에게르텔라와 아시네토박터가 낮은 빈도를 보였다.

[0094] 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체에 있어서 프라이머에 따라 박테리아 빈도에 큰 차이를 나타내지 않았다. 다른 프라이머를 이용하여 증폭된 개개인의 샘플 별 인덱스는 어떤 중요도도 보이지 않았다; V1V2, V3V4 및 V6V8 프라이머 각각에서, 게놈 DNA 인공 유전체의 경우 0.139, 0.179 및 0.193이고, 플라스미드 인공 유전체의 경우 0.224, 0.263 및 0.179이며, PCR 인공 유전체의 경우 0.158, 0.225 및 0.203의 결과를 보였다.

[0095] 각 프라이머 세트들의 차이를 비교하기 위하여, 도 5의 히트맵(heatmap)과 계통도(dendrogram)를 통해 속 수준에서 분석된 인공 유전체의 박테리아 빈도를 나타내었다. 상기 계통도의 경우 V6V8 영역을 타겟하는 인공 유전체는 함께 클러스터링된 반면, V1V2 또는 V3V4 영역을 타겟하는 경우 명확한 클러스터링을 보이지 않았다. 3개의 16S rRNA 유전자 영역(V1V2, V3V4 및 V6V8) 모두에서 플라스미드 인공 유전체와 PCR 인공 유전체는 게놈 DNA 인공 유전체 보다 근접한 결과를 보였다.

[0096] 도 6에는 인공 유전체 결과 값과 예측되는 결과 비율을 이용하여 Bray-Curtis 인덱스를 이용해 PCA를 나타내었다. 예측되는 결과 비율은 투입(input) 농도와 16S rRNA 유전자 복제수를 곱한 뒤 게놈 사이즈와 16S rRNA 유전자의 GC 함량으로 나누어서 계산하였다. 인공 유전체에서 예측되는 비율은 분홍색 별 모양으로 표시하였다. 각 모양은 타겟하는 16S rRNA 유전자 프라이머를 나타내며, 색깔은 인공 유전체를 나타내도록 하였다. 도 6의 그래프에서 녹색과 청색 군은 함께 모여 분포하였으나, 적색 군은 이들과 떨어져 존재하였다. 예측되는 비율은 인공 유전체 샘플 어느 것과도 함께 분포하지 않으나, V3V4 영역에서 증폭된 게놈 DNA 인공 유전체에서는 근접하게 분포된 것을 볼 수 있었다.

[0098] (3) 세포 인공 유전체를 이용한 DNA 추출 방법과 프라이머의 영향 평가

[0099] 세포 인공 유전체는 18종의 박테리아 세포를 OD600 값 동량으로 포함하여 제조되었다. 이들 DNA는 4종류의 상업적으로 판매되고 있는 키트로, MP Bio soil kit (MP), Qiagen blood and tissue kit (QiaB), Qiagen stool mini kit (QiaS) 및 Sigma GenElute bacterial DNA kit (Sig)를 이용하여 추출하였다. 상이한 키트를 사용하여 추출된 DNA를 3개의 프라이머 세트를 사용하여 증폭하였다. QiaB를 사용하여 추출된 DNA는 추가의 실험을 수행할 수 있을 정도의 충분한 양으로 증폭시키지 못하여 제외하였다.

[0100] 문 수준에서 분석한 세포 인공 유전체는 각 추출 방법에 따라 다양한 박테리아 비율을 보였다(도 7). 세포수만을 고려하였을 때 세포 인공 유전체는 프로테오박테리아 50.8%, 피르미쿠테스 24.6%, 악티노박테리아 15.9% 및 박테로이데스 8.7%에 해당하지만(청색 점선), 16S rRNA 유전자의 16S rRNA 복제수, 게놈 사이즈 및 GC 함량을 고려한다면, 예측되는 비율이 프로테오박테리아 35.8%, 피르미쿠테스 42.3%, 악티노박테리아 14.1% 및 박테로이데스 7.7%에 해당하였다(녹색 점선). 세포 mock DNA에 있어서, 모든 DNA 추출 방법에서 동일하게 총 박테리아 중 프로테오박테리아가 가장 높은 빈도를 보였다. MP 및 Sig에서 2번째로 높은 빈도를 보인 문은 피르미쿠테스에 해당하였고, 그 다음으로 박테로이데스가 악티노박테리아 보다 높은 빈도를 나타냈다. 하지만, QiaS 추출 방법은 악티노박테리아가 가장 낮은 빈도를 보였으며, 박테로이데스가 피르미쿠테스 보다 높은 빈도를 보였다. QiaB 추출 방법에서는 V1V2 및 V3V4 영역 각각에서 프로테오박테리아 85.7% 및 75.7%, 박테로이데스 9.4% 및 15.2%, 피르미쿠테스 4.1% 및 7.5%, 및 악티노박테리아 0.8% 및 1.7%의 분포로 측정된 것을 확인할 수 있었다.

[0101] 종 수준에 있어서는 도 8에서 보는 바와 같이 3종류의 프라이머 세트를 사용하여 시퀀싱된 DNA 추출 방법 모두에 있어서 예측되는 박테리아 비율을 보이는 것은 없었다. 큰 청색 바는 세포수, 16S rRNA 유전자 복제 수, 게놈 사이즈 및 16S rRNA 유전자의 GC 함량을 고려한 세포 인공 유전체에서 예측되는 비율을 나타낸 것이다. 세포수만을 고려하여 예측되는 비율은 녹색 바로 나타내었다. 박테로이데스 및 캠필로박터 속에서 QiaS로 추출된 인공 유전체에서 박테리아 빈도는 20%를 초과하였고, 이는 캠필로박터 속에서 QiaB 방법으로 추출된 경우 또한 20%를 초과하는 값을 나타내었다. 각 추출 방법에 따른 차이를 측정하기 위하여, Bray-Curtis 인덱스를 사용하여 히트맵과 계통도를 나타내었다(도 9).

[0102] 18종의 박테리아의 퍼센티지에서 박테리아 세포수에 의해 계산된 예측되는 빈도 비율을 계산하였다. 그 결과, 모든 추출 방법에서 박테로이데스와 캠필로박터의 빈도가 매우 높은 수준으로 관찰되었다. 단, 캠필로박터 속에

서 V6V8 프라이머를 사용한 경우는 V1V2 프라이머나 V3V4 프라이머를 사용한 경우와 비교하여 빈도가 상대적으로 낮게 나타났다. 세포 인공 유전체에 있어서 DNA 추출 방법은 프라이머와 달리 조밀하게 분포한 것을 확인할 수 있었다. V1V2 및 V3V4 프라이머를 사용하여 시퀀싱된 세포 mock DNA는 상대적으로 V6V8 프라이머 보다 가깝게 분포하는 것을 확인할 수 있었다.

[0103] 한편, PCA는 상이한 프라이머를 이용하여 증폭된 세포 인공 유전체에 있어서 DNA 추출 방법 사이의 거리를 나타내었다(도 10). 프라이머의 종류 보다 DNA 추출 방법에 따라 군집이 형성된 것을 볼 수 있었다. 별 모양은 예측되는 비율을 나타내고, 청색 및 분홍색은 각각 이론 비율과 세포수를 나타낸다.

[0105] (4) 인공 유전체에서 차세대 염기서열 분석의 결과물에 영향을 미치는 박테리아 특징

[0106] 차세대 염기서열 분석의 결과물에 어떠한 박테리아 특성이 영향을 미치는 지 분석하기 위하여, 공정에 영향을 미칠 수 있는 박테리아 요소로, 차세대 염기서열 분석 시 투입한 각 박테리아 별 시료 양(input), 부피(volume), 16S rRNA 유전자 복제 수, 게놈 사이즈 및 시퀀스의 GC 함량을 선별하였고, 이를 이용하여 이론적 빈도로 하기 식 3을 도출할 수 있었다:

[0108] [식 3]

[0109] 박테리아 빈도 = 투입 농도(input concentration) X 16S rRNA 유전자 복제 수 X 게놈 사이즈 X ((-1.2) X V3V4 서열의 %GC 함량)

[0110] 박테리아 특성은 게놈 DNA와 관련이 있기 때문에, 도 10에서는 게놈 DNA 인공 유전체만을 고려하였다. 도 11의 그래프에서 선은 각각의 박테리아의 예측 빈도를 나타내며, 색깔은 예측 선에서 고려된 요소를 나타낸 것이다. 적색 선은 투입 농도; 주황색 선은 투입 농도 및 16S rRNA 유전자 복제 수; 황색 선은 투입 농도, 16S rRNA 유전자 복제 수 및 게놈 사이즈; 녹색 선은 투입 농도 및 V3V4 서열의 %GC 함량을 반영한 것이다. 또한, 이론적 비율을 나타낸 청색 선은 투입 농도, 16S rRNA 유전자 복제 수, 게놈 사이즈 및 V3V4 서열의 %GC 함량을 측정한 것이고, 이는 게놈 DNA 인공 유전체에서 박테리아 빈도 결과 값과 가장 근접한 것을 볼 수 있었다.

[0111] 보다 정밀한 분석을 위하여, SPSS를 이용하여 세포 인공 유전체, 게놈 DNA 인공 유전체, PCR 인공 유전체 및 플라스미드 인공 유전체 각각에 있어서, 박테리아의 특성을 고려하여 각기 예측되는 박테리아의 빈도와, V3V4 영역에서 증폭된 게놈 DNA 인공 유전체의 실제 결과 값에 대하여 다중 회귀 모형(multiple regression model)을 분석해 그 결과를 표 5 내지 8에 나타내었다.

표 5

[0113]

구분	Coefficient		
	최소	최대	평균
상수	39.063	39.963	39.522
투입 농도(세포수)	6.558E-08	7.276E-08	6.905E-08
V3V4 %GC 함량(%)	-0.444	-0.418	-0.431
16S rRNA 유전자 복제 수(개수)	0.044	0.114	0.075
게놈 사이즈(bp)	-2.411E-06	-2.286E-06	-2.334E-06
그람 양성 여부	-7.429	-7.059	-7.253

표 6

[0115]

구분	Coefficient			Standard Error	t-Statistic	P-value
	최소	최대	평균			
상수	16.43	18.557	17.327	8.796	2.249	0.044
투입 농도(세포수)	0.471	0.555	0.53375	0.395	1.499	0.173
V3V4 %GC 함량(%)	-0.431	-0.343	-0.388	0.142	-3.291	0.006
16S rRNA 유전자 복제 수(개수)	0.415	0.515	0.46125	0.173	3.638	0.003
게놈 사이즈(bp)	-8.292E-07	-4.816E-07	-6.17075E-07	0	-2.055	0.062
그람 양성 여부	-0.731	-0.34	-0.49775	1.821	-0.184	0.857
R-squared			0.702			
Adjusted R-squared			0.578			

표 7

[0117]

구분	Coefficient			Standard Error	t-Statistic	P-value
	최소	최대	평균			
상수	14.188	26.537	20.3895	11.026	1.849	0.122
투입 농도(㎕)	0.192	0.288	0.25225	0.400	0.632	0.539
V3V4 %GC 함량(%)	-0.484	-0.237	-0.365	0.149	-2.395	0.036
R-squared			0.292			
Adjusted R-squared			0.197			

표 8

[0119]

구분	Coefficient			Standard Error	t-Statistic	P-value
	최소	최대	평균			
상수	12.44	23.635	17.63375	9.203	1.893	0.095
투입 농도(㎕)	0.223	0.669	0.4565	0.319	1.441	0.246
V3V4 %GC 함량(%)	-0.421	-0.371	-0.399	0.133	-3.005	0.009
R-squared			0.414			
Adjusted R-squared			0.336			

[0121] 상기 표 5 내지 8의 결과를 토대로 게놈 DNA 인공 유전체, 세포 인공 유전체, PCR 인공 유전체 및 플라스미드 인공 유전체 각각에서 박테리아의 예측 빈도를 도출하는 수식으로 하기 식 2 및 4 내지 6을 도출할 수 있었다. 또한, 상기 게놈 DNA 인공 유전체에서 하기 식 2에 대입할, 18종의 박테리아 균주 각각에 대하여 측정된 시료양(input), V3V4 영역의 GC 함량(%), 16S rRNA 유전자 복제 수, 게놈 사이즈(bp) 및 그람 양성 여부를 표 9에 나타내었고, 이들을 하기 식 4에 대입하여 예측되는 각 박테리아별 빈도를 표 10에 나타내었으며, 이러한 예측 빈도를 그 박테리아의 실제 빈도를 비교한 결과를 도 12(A)에 나타내었다. 또한, 상기 세포 인공 유전체에서 하기 식 4에 대입할, 18종의 박테리아 균주 각각에 대하여 측정된 세포 수(input), V3V4 영역의 GC 함량(%), 16S

rRNA 유전자 복제 수, 게놈 사이즈(bp) 및 그람 양성 여부를 표 11에 나타내었다. 또한, PCR 인공 유전체 및 플라스미드 인공 유전체 각각에서 하기 식 5 및 6에 의해 예측되는 각 박테리아별 빈도와 그 박테리아의 실제 빈도를 비교한 결과를 도 12(B) 및 12(C)에 나타내었다. 그 결과, 각 인공 유전체에서 예측되는 각 박테리아의 빈도가 실제 결과 값과 상당히 유사한 패턴을 갖는 것을 알 수 있었다.

[0123] [식 2]

[0124] 목표 박테리아의 예측 분포 비율(%) = A1 + A2 X (목표 박테리아의 시료의 양(㎕)) + A3 X (V3V4 영역의 GC 함량(%)) + A4 X (16S rRNA 유전자 복제 수(개수)) + A5 X (게놈 사이즈(bp)) + A6 X (그람 양성 여부)

[0125] 상기 식 2에서, A1은 17.327이고, A2는 0.53375이며, A3는 -0.388이고, A4는 0.46125이며, A5는 -6.17075E-07이고, b6는 -0.49775이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0이다.

[0127] [식 4]

[0128] 목표 박테리아의 예측 분포 비율(%) = a1 + a2 X (목표 박테리아의 수(세포수)) + a3 X (V3V4 영역의 GC 함량(%)) + a4 X (16S rRNA 유전자 복제 수(개수)) + a5 X (게놈 사이즈(bp)) + a6 X (그람 양성 여부)

[0129] 상기 식 2에서, 상기 a1은 39.522이고, a2는 6.905E-08이며, a3는 -0.431이고, a4는 0.075이며, a5는 -2.334E-06이고, a6는 -7.253이며, 상기 '그람 양성 여부'는 목표 박테리아가 그람 양성인 경우 1이고, 그람 음성인 경우 0이다.

[0131] [식 5]

[0132] 목표 박테리아의 분포 비율(%) = c1 + c2 X (시료의 양) + c3 X (V3V4 영역의 GC 함량)

[0133] 상기 식 5에서, c1은 20.3895이고, c2는 0.25225이며, c3는 -0.365이다.

[0135] [식 6]

[0136] 목표 박테리아의 분포 비율(%) = d1 + d2 X (시료의 양) + d3 X (V3V4 영역의 GC 함량)

[0137] 상기 식 6에서, d1은 17.63375이고, d2는 0.4565이며, d3는 -0.399이다.

표 9

[0139]

구분	시료 양 (ul)	V3V4 영역 GC 함량(%)	16S rRNA 복제수 (갯수)	게놈 사이즈 (bp)	그람 양성 여부 0=neg, 1=pos
Acinetobacter	22	51.2	6	4,028,903	0
Actinomyces	22	59.1	3	2,393,958	1
Aeromonas	20	54.1	10	4,744,448	0
Bacillus	20	52.5	13	5,427,083	0
Bacteroides	20	46.7	6	5,241,700	0
Bifidobacterium	21	58.6	5	2,089,645	1
Campylobacter	22	51.4	3	1,766,442	1
Clostridium	22	53.4	11	4,207,674	1
Delftia	22	53.2	5	6,953,182	0
Eggerthella	21	60.5	3	3,632,260	1
Enterobacter	21	53	4	2,881,400	1
Enterococcus	19	54.7	7	5,037,933	0

Escherichia	22	56	8	5,470,076	0
Klebsiella	18	50.4	5	1,867,005	1
Lactobacillus	20	51.7	4	6,073,945	0
Pseudomonas	22	56	8	5,598,796	0
Staphylococcus	20	50.9	5	2,761,522	1
Streptococcus	20	52.8	4.5	2,110,494	1

표 10

[0141]

Bacteria Composing Mock	Standard (%)	Range (%)	Average (%)	Minimum (%)	Maximum (%)
Acinetobacter	6.4	5.4 - 7.6	6.435514681	5.422880315	7.561033632
Actinomyces	3.6	3.1 - 4.1	3.606698367	3.184796906	4.087496709
Aeromonas	6.8	6.5 - 7.0	6.75526975	6.511461005	6.987673843
Bacillus	8.3	8.1 - 8.6	8.338582758	8.116803343	8.596316827
Bacteroides	7.8	7.3 - 8.1	7.751877973	7.38488236	8.09379728
Bifidobacterium	4.1	4.0 - 4.2	4.099982312	4.04109149	4.178666366
Campylobacter	7.5	6.8 - 8.2	7.536022803	6.860557134	8.153986251
Clostridium	8.4	8.0 - 8.8	8.389099566	8.077794297	8.718984202
Delftia	3.4	2.8 - 4.1	3.393765217	2.849547549	4.098561108
Eggerthella	1.0	0.6 - 1.4	0.933873161	0.610703584	1.309232926
Enterobacter	5.0	4.6 - 5.5	5.045710095	4.63831776	5.43897114
Enterococcus	5.3	4.0 - 6.8	5.255617494	4.065845956	6.370931467
Escherichia	5.7	5.2 - 6.2	5.715302852	5.256647228	6.159611398
Klebsiella	5.8	5.3 - 6.3	5.81771789	5.391443287	6.250698426
Lactobacillus	3.8	3.5 - 4.0	3.821320389	3.509784806	3.964135243
Pseudomonas	5.6	5.1 - 6.1	5.635872958	5.1723485	6.097619846
Staphylococcus	5.9	5.3 - 6.4	5.861983812	5.325359043	6.394879242
Streptococcus	5.3	4.7 - 5.9	5.295891915	4.75488609	5.848678375

표 11

[0143]

구분	세포 수 Cell count no.	V3V4 영역 GC 함량(%)	16S rRNA 복제수 (갯수)	게놈 사이즈 (bp)	그람 양성 여부 0=neg, 1=pos
Acinetobacter	20320000	51.2	6	4,028,903	0
Actinomyces	26080000	59.1	3	2,393,958	1
Aeromonas	5600000	54.1	10	4,744,448	0
Bacillus	9600000	52.5	13	5,427,083	0
Bacteroides	37600000	46.7	6	5,241,700	0
Bifidobacterium	21760000	58.6	5	2,089,645	1
Campylobacter	52480000	51.4	3	1,766,442	1
Clostridium	15840000	53.4	11	4,207,674	1
Delftia	11520000	53.2	5	6,953,182	0
Eggerthella	20640000	60.5	3	3,632,260	1
Enterobacter	18080000	53	4	2,881,400	1
Enterococcus	20480000	54.7	7	5,037,933	0
Escherichia	16320000	56	8	5,470,076	0

Klebsiella	28480000	50.4	5	1,867,005	1
Lactobacillus	22400000	51.7	4	6,073,945	0
Pseudomonas	65600000	56	8	5,598,796	0
Staphylococcus	17600000	50.9	5	2,761,522	1
Streptococcus	19840000	52.8	4.5	2,110,494	1

[0145] (5) gDNA mock, 플라스미드 mock 및 PCR mock에 있어서, 각 박테리아 종의 투입 농도의 차세대 염기서열 분석 결과에 미치는 영향

[0146] 각기 다른 비율에서 3종류의 인공 유전체는 정량적으로 회석된 비율을 보였다(도 13). 게놈 DNA 인공 유전체에서 그룹 A와 그룹 B의 1:1 비율은 각각 62.7% 및 37.3%로 측정되었다. 또한, 플라스미드 인공 유전체의 경우 그룹 A와 그룹 B를 1:1 비율로 혼합한 경우 각각 50.2% 및 49.8%를 보였고, PCR 인공 유전체의 경우 동일 비율에서 52.7% 및 47.3%로 측정되었다. 비록 gDNA 인공 유전체는 그룹 A와 그룹 B가 1:1 비율을 보이지 않았지만, 1:2, 1:4, 1:10 및 1:100의 비율로 혼합하였을 때 비례적 정량 변화를 관찰할 수 있었다. 플라스미드 인공 유전체와 PCR 인공 유전체는 혼합 비율에 따라 비례적 변화를 보였다. 단, 게놈 DNA 인공 유전체, 플라스미드 인공 유전체 및 PCR 인공 유전체 모두에 있어서, 1:10 및 1:100 비율로 혼합하였을 때 그룹 A 내 박테리아 빈도는 1% 컷오프 값을 넘지 않았다.

[0148] (6) 생물정보학 분석 플랫폼과 데이터베이스의 영향

[0149] 인공 유전체를 도 14와 같이 Mothur and Illumina; BaseSpace with Silva; Eztaxon 및 Greengenes 데이터베이스로 분석하였다. 게놈 DNA 인공 유전체 (도 14A), 플라스미드 인공 유전체 (도 14B) 및 PCR 인공 유전체 (도 14C)은 모두 상기 3종의 데이터베이스로, Mothur-Eztaxon (청색), Mothur-Silva (분홍색) and BaseSpace-Greengenes(회색) 모두 유사한 경향을 보였다. Greengenes 데이터베이스를 이용한 Illumina BaseSpace는 에어로모나스(Aeromonas) 및 클로스트리디움 속에서 Silva를 이용한 Mothur 또는 Eztaxon을 이용한 Mothur 보다 다소 논란의 소지가 많은 결과를 보였다. 속 수준에 있어서 BaseSpace로 분석한 경우 18종 박테리아에 포함되지 않은 플레시오모나스(Plesiomonas)와 알칼리필러스(Alkaliphilus)가 검출되었다. Silva 데이터베이스를 이용한 Mothur로 분석한 데이터의 경우 Mothur 웹사이트에서 제공하는 표준 프로토콜(protocol)에 따라 수행하였음에도 불구하고 엔테로박터(Enterobacter)를 검출하지 못하였고, 이는 살모넬라(Salmonella)와 엔테로박터리아시아-미분류(Enterobacteriaceae_unclassified)로 분류되었다. 다만, Eztaxon 데이터베이스를 이용한 Mothur로부터 얻어진 데이터에서는 인공 유전체의 18 속이 모두 규명 되었다.

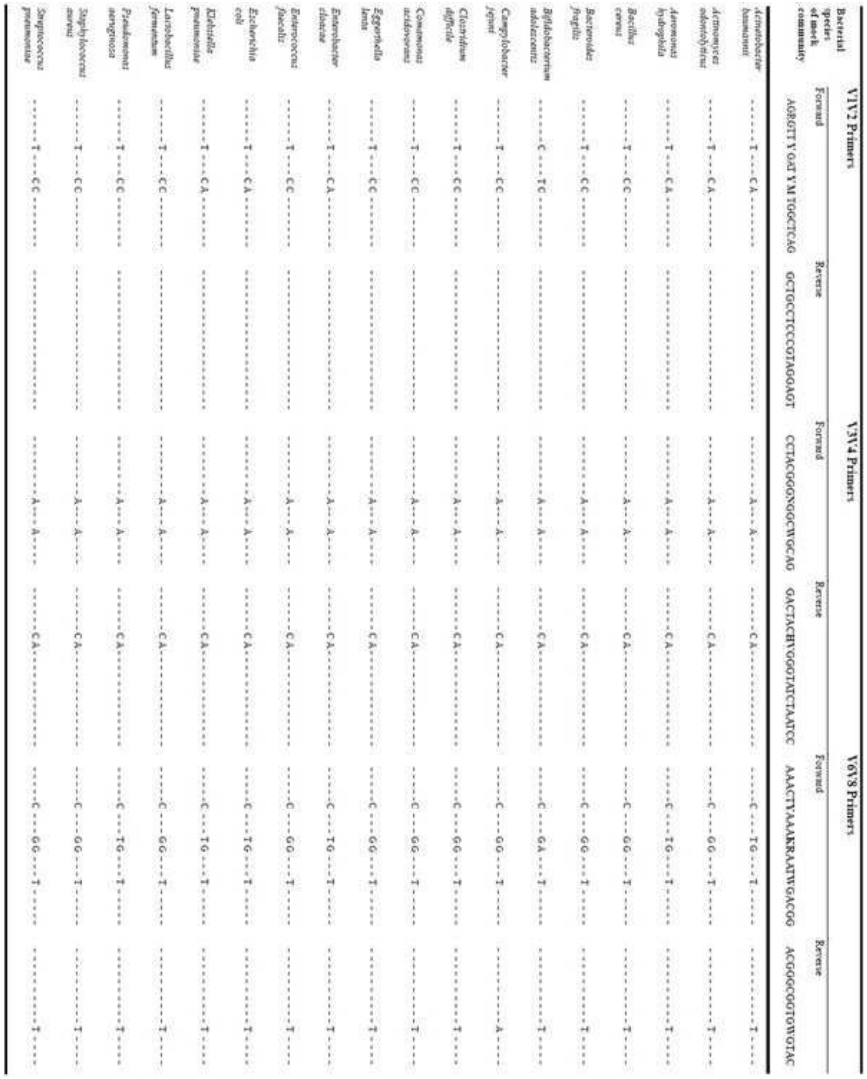
[0150] 본 실험에서 사용된 3가지 파이프라인을 통해, Eztaxon을 종 수준에서 분류 체계(taxonomic classifications)로 선택하였다. 이후, Eztaxon의 종 분석을 평가하였고, 사용되는 프라이머에 따라 몇 종이 잘못 분석된 것을 확인할 수 있었다(표 5). 모든 프라이머에 있어서 에어로모나스 하이드로필라(Aeromonas hydrophila)는 검출되지 않았고, V1V2, V3V4 및 V6V8 프라이머 각각에서 에어로모나스 타이와넨시스(A. taiwanensis), 에어로모나스 몰루스코룸(A. molluscorum), 에어로모나스 미디아(A. media)로 분석되었다.

도면

도면1

Bacteria	Type Strain No.	Genome Size	16S rRNA gene Copy No.	Genome %GC Content	V1V2 %GC Content	V3V4 %GC Content	V8V9 %GC Content	Gram-staining
Acinetobacter baumannii	ATCC 19606	4,028,903	6	54.6	55.6	51.2	52.7	Negative
Actinomyces odontolyticus	ATCC 17929	2,393,958	3	65.4	57.6	59.1	56.2	Positive
Aeromonas hydrophila	ATCC 7966	4,744,448	10	35.3	57.7	54.1	55.8	Negative
Bacillus cereus	ATCC 14579	5,427,083	13	35.8	54.7	52.5	53.2	Positive
Bacteroides fragilis	ATCC 25285	5,241,700	6	43.2	48.7	46.7	53.5	Negative
Bifidobacterium adolescentis	ATCC 15703	2,089,645	5	59.2	60.3	58.6	61.4	Positive
Campylobacter jejuni	ATCC 33560	1,766,442	3	51.8	46.1	51.4	49.5	Negative
Clostridium difficile	ATCC 9689	4,207,674	11	39.1	50.8	53.4	50.9	Positive
Dolfsia acidaminophila	ATCC 15668	6,953,182	5	39.5	56.3	53.2	53.5	Negative
Eggerthella lenta	ATCC 25559	3,632,360	3	57	63.8	60.5	62.0	Positive
Enterobacter cloacae	ATCC 13047	5,598,706	8	60.1	56.2	56	53.8	Negative
Enterococcus faecalis	ATCC 19433	2,881,400	4	32.8	55.8	53	53.4	Positive
Escherichia coli	ATCC 11775	5,037,993	7	64.2	55.7	54.7	54.2	Negative
Klebsiella pneumoniae	ATCC 13883	5,470,076	8	38.4	57.1	56	52.5	Negative
Lactobacillus fermentum	ATCC 14931	1,867,005	5	61.5	52.5	50.4	55.4	Positive
Pseudomonas aeruginosa	ATCC 10145	6,073,845	4	43.1	55.5	51.7	54.5	Negative
Staphylococcus aureus	ATCC 12600	2,761,522	5	50.6	51.1	50.9	49.6	Positive
Streptococcus pneumoniae	ATCC 33400	2,221,315	4	39.6	52.9	51.5	53.7	Positive

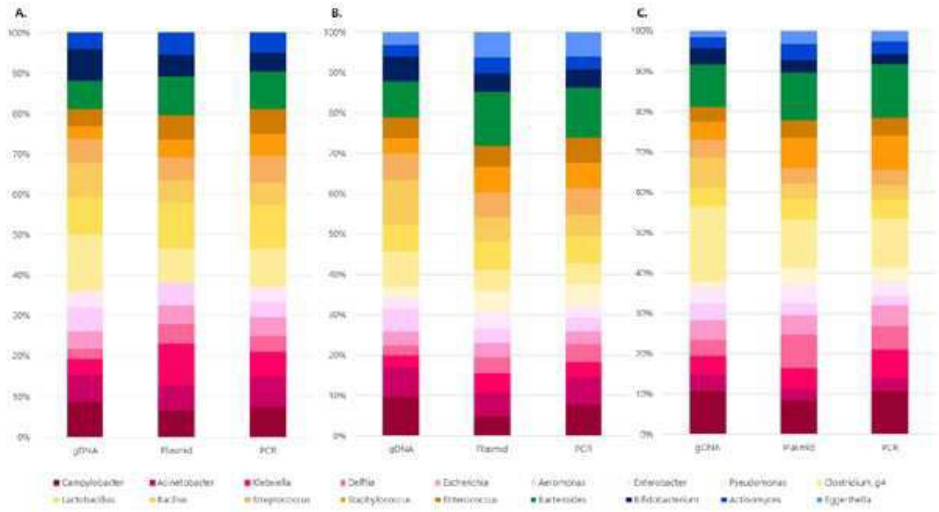
도면2



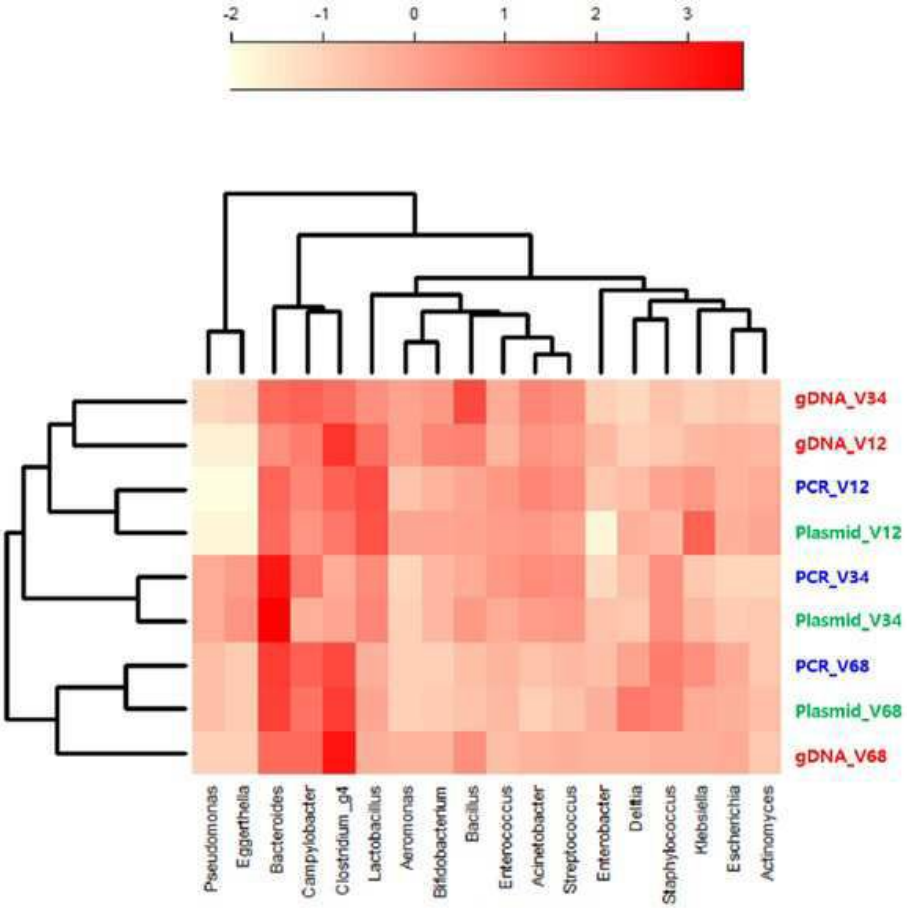
Bacterial species of mock community	V1V2 Primers		V3V4 Primers		V6V8 Primers	
	Forward	Reverse	Forward	Reverse	Forward	Reverse
	AGRGTTYGATYMTGGCTCAG	GCTGCCTCCCGTAGGAGT	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	AAACTYAAAKRAATWGACGG	ACGGGCGGTGTWTRCAC
Acinetobacter baumannii	------T---CA--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Actinomyces odontolyticus	------T---CA--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Akkermansia muciniphila	------T---CA--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Bacillus cereus	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Bacteroides fragilis	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Bifidobacterium adolescentis	------C---TC--------	----------	------A---A-----	------CA----------	------C---GA---T----	----------T-----
Campylobacter jejuni	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------A-----
Clostridium difficile	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Cutibacterium acnes	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Eggerthella lenta	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Enterobacter cloacae	------T---CA--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Enterococcus faecalis	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Escherichia coli	------T---CA--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Klebsiella pneumoniae	------T---CA--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Lactobacillus fermentum	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Pseudomonas aeruginosa	------T---CC--------	----------	------A---A-----	------CA----------	------C---TG---T----	----------T-----
Staphylococcus aureus	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----
Streptococcus pneumoniae	------T---CC--------	----------	------A---A-----	------CA----------	------C---GG---T----	----------T-----

도면3

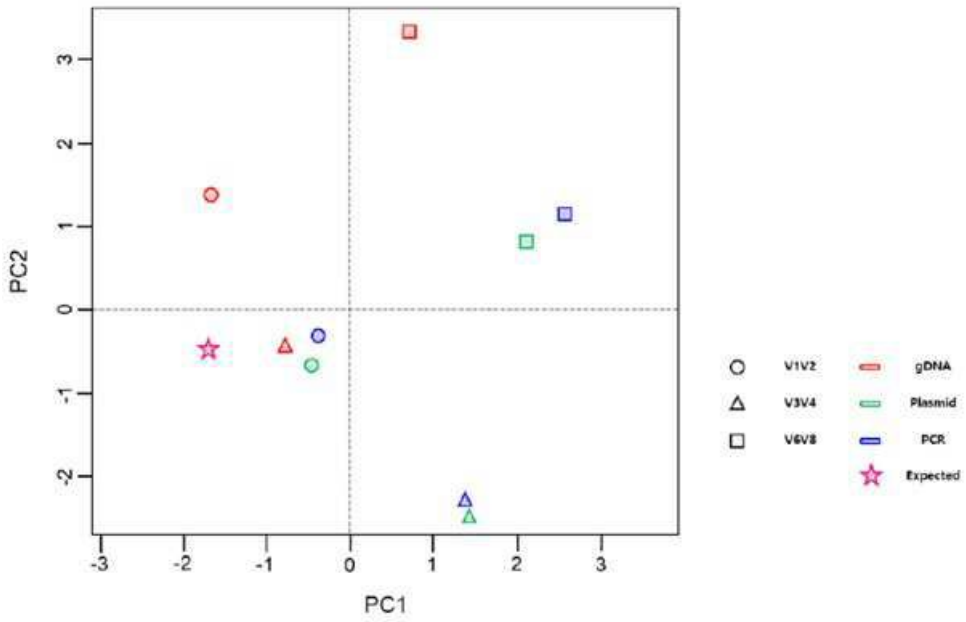
도면4


A.
B.
C.
gDNA
Plasmid
PCR

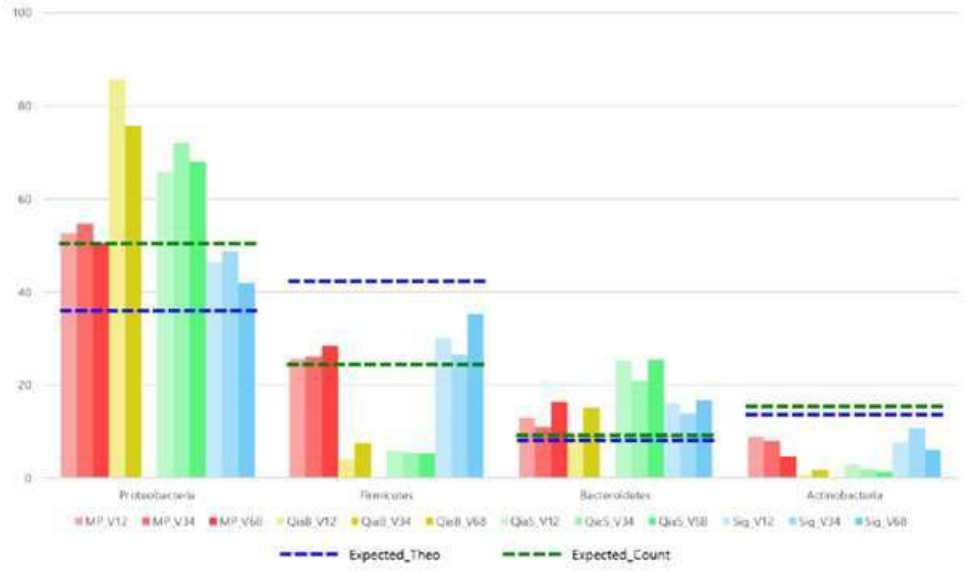
도면5


-2
-1
0
1
2
3
gDNA_V34
gDNA_V12
PCR_V12
Plasmid_V12
PCR_V34
Plasmid_V34
PCR_V68
Plasmid_V68
gDNA_V68
Pseudomonas
Eggerthella
Bacteroides
Campylobacter
Clostridium_g4
Lactobacillus
Aeromonas
Bifidobacterium
Bacillus
Enterococcus
Acinetobacter
Streptococcus
Enterobacter
Delftia
Staphylococcus
Klebsiella
Escherichia
Actinomyces

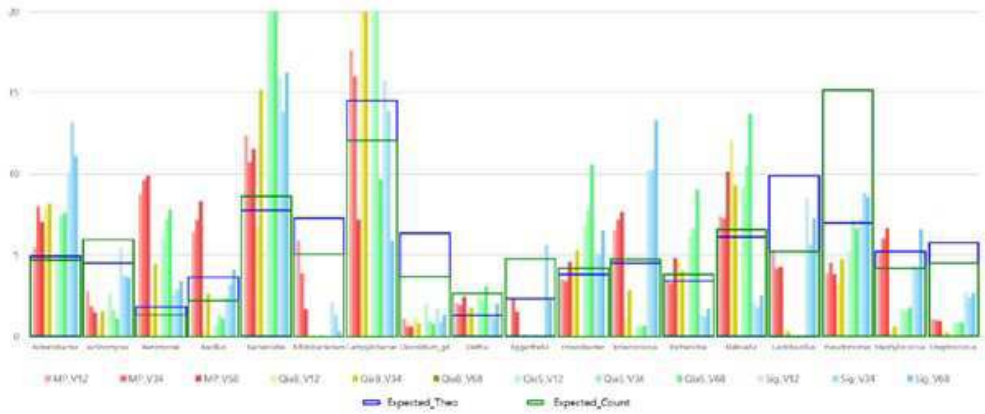
도면6


PC2
PC1
V1V2
V3V4
V6V8
gDNA
Plasmid
PCR
Expected

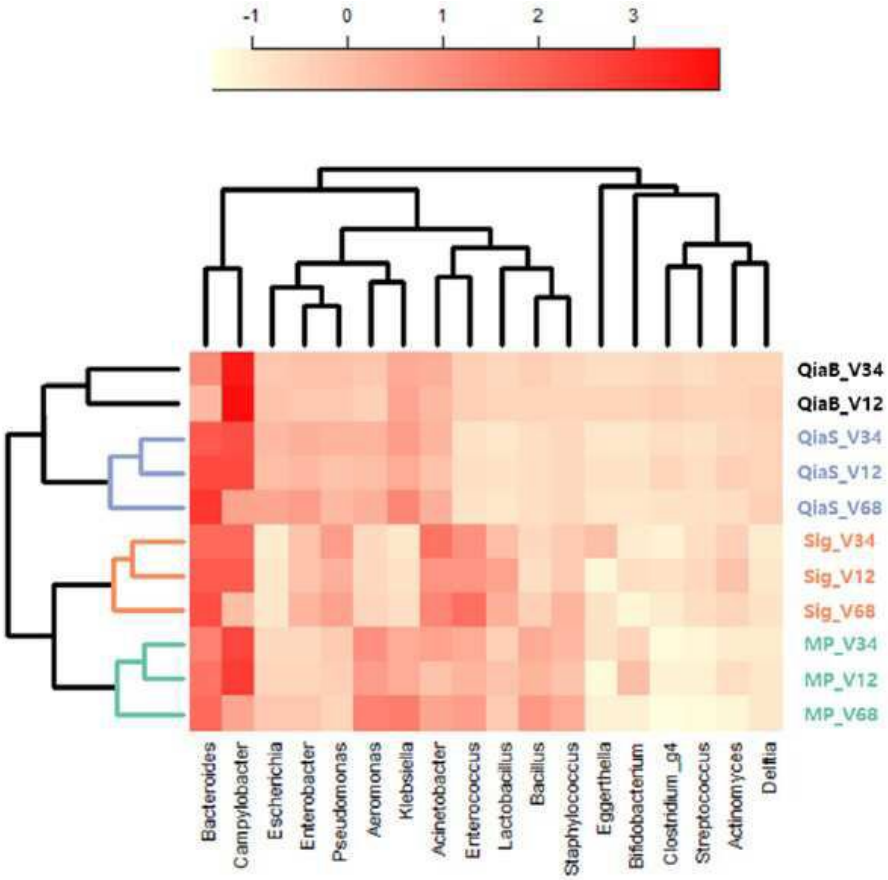
도면7


Proteobacteria
Firmicutes
Bacteroidetes
Actinobacteria
Expected_Theo
Expected_Count

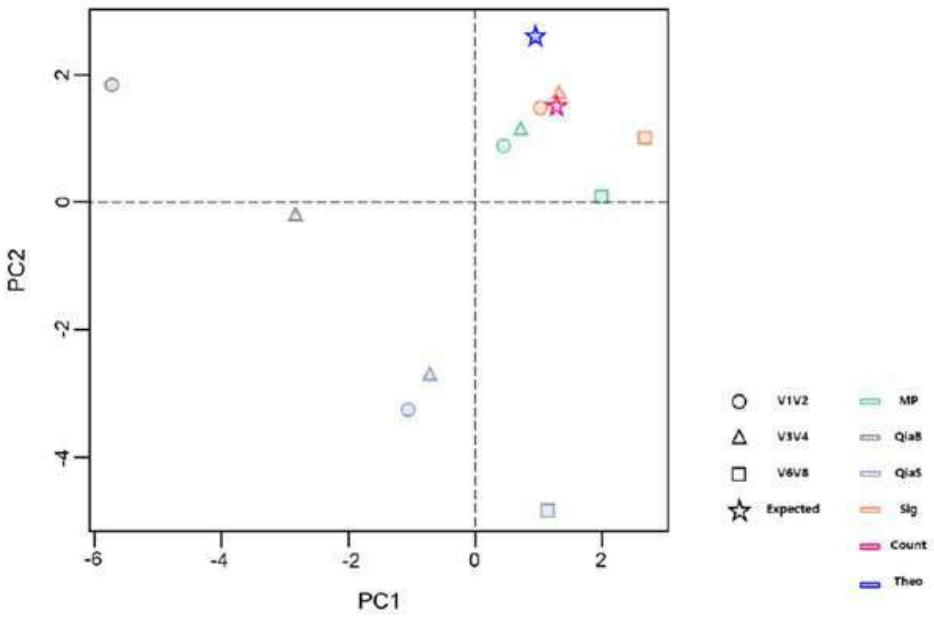
도면8


Expected_Theo
Expected_Count

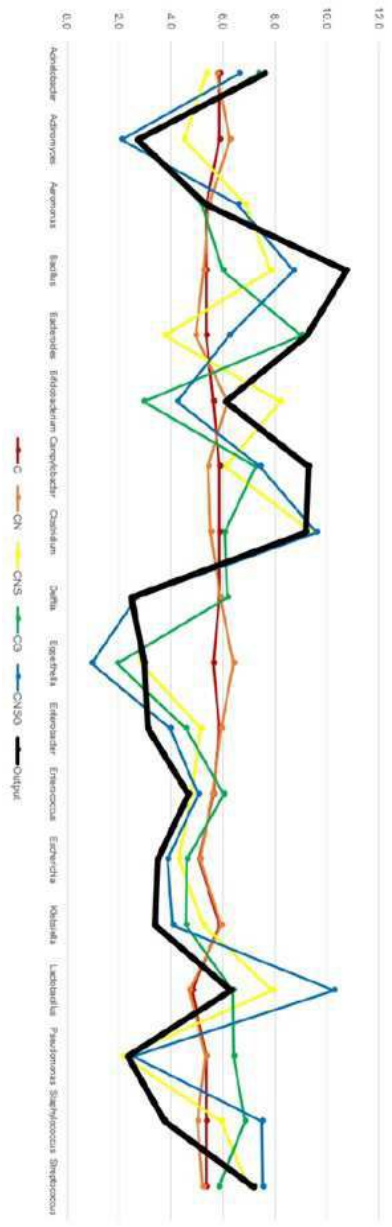
도면9


-1
0
1
2
3
QiaB_V34
QiaB_V12
QiaS_V34
QiaS_V12
QiaS_V68
Sig_V34
Sig_V12
Sig_V68
MP_V34
MP_V12
MP_V68
Bacteroides
Campylobacter
Escherichia
Enterobacter
Pseudomonas
Aeromonas
Klebsiella
Acinetobacter
Enterococcus
Lactobacillus
Bacillus
Staphylococcus
Eggerthella
Bifidobacterium
Clostridium_g4
Streptococcus
Actinomyces
Delftia

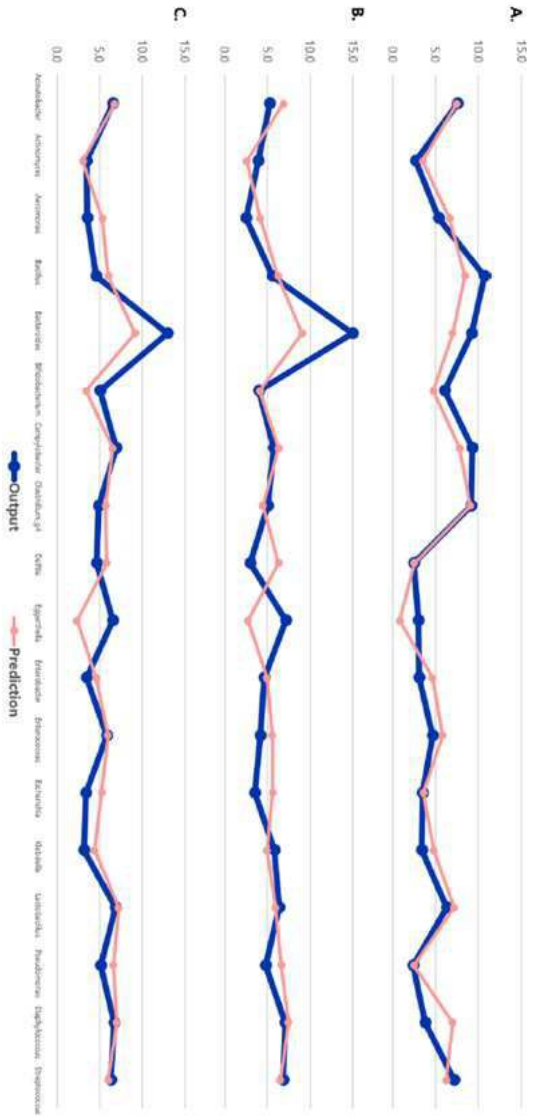
도면10


PC2
PC1
2
0
-2
-4
-6
-4
-2
0
2
V1V2
V3V4
V6V8
Expected
MP
QiaB
QiaS
Sig
Count
Theo

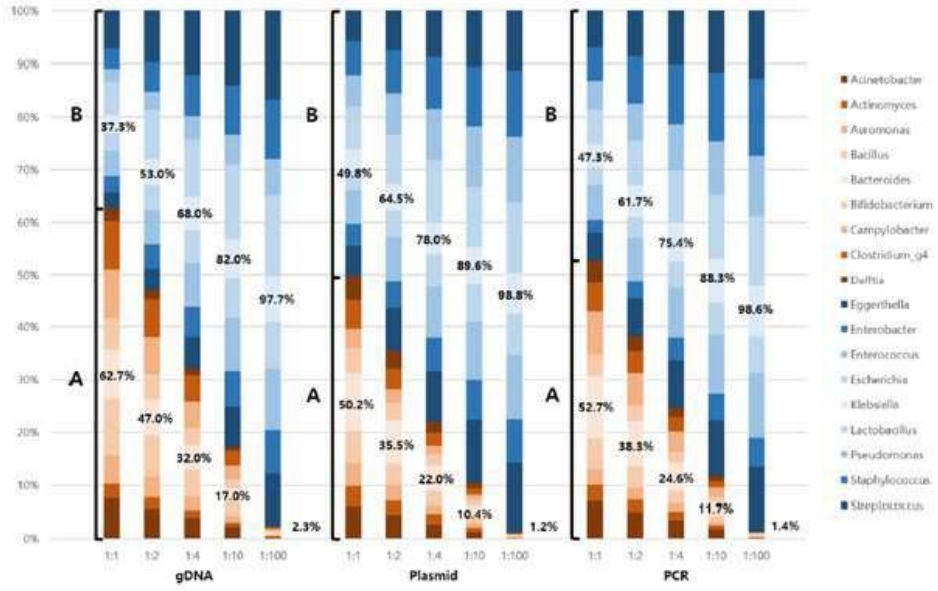
도면11


0.0
2.0
4.0
6.0
8.0
10.0
12.0
Achelobacter
Adlercreutzia
Agromonas
Bacillus
Bacteroides
Bifidobacterium
Campylobacter
Clostridium
Delftia
Egerthella
Enterobacter
Enterococcus
Escherichia
Klebsiella
Lactobacillus
Pseudomonas
Staphylococcus
Streptococcus
C
CN
CNS
CO
CNSO
Output

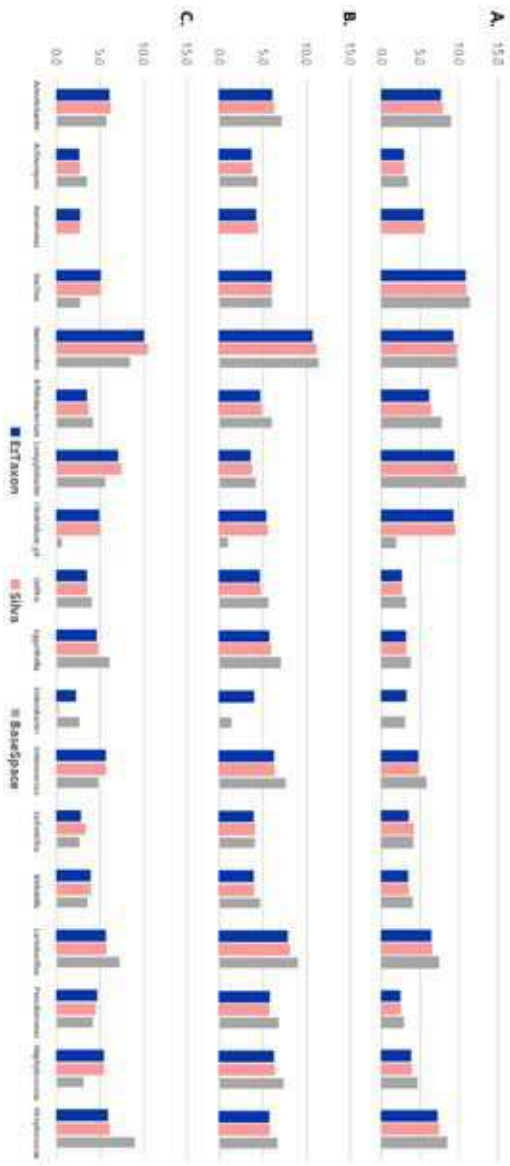
도면12


A.
B.
C.
0.0
5.0
10.0
15.0
Output
Prediction

도면13


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A
B
37.3%
62.7%
53.0%
47.0%
68.0%
32.0%
82.0%
17.0%
97.7%
2.3%
49.8%
50.2%
64.5%
35.5%
78.0%
22.0%
89.6%
10.4%
98.8%
1.2%
47.3%
52.7%
61.7%
38.3%
75.4%
24.6%
88.3%
11.7%
98.6%
1.4%
1:1
1:2
1:4
1:10
1:100
gDNA
Plasmid
PCR
Acinetobacter
Actinomyces
Aeromonas
Bacillus
Bacteroides
Bifidobacterium
Campylobacter
Clostridium_g4
Delftia
Eggerthella
Enterobacter
Enterococcus
Escherichia
Klebsiella
Lactobacillus
Pseudomonas
Staphylococcus
Streptococcus

도면14



도면15

Bacteria composing mock	V1V2	V3V4	V6V8
Acinetobacter baumannii	○	○	○
Actinomyces odontolyticus	○	○	○
Aeromonas hydrophila	*A. taiwanensis*	*A. molluscorum*	*A. media*
Bacillus cereus	○	*B. bingmayongensis*	*B. toyonensis*
Bacteroides fragilis	○	○	○
Bifidobacterium adolescentis	○	*B. stercoris*	○
Campylobacter jejuni	*C. coli*	○	*C. lari*
Clostridium difficile	○	○	○
Delftia acidovorans	○	○	○
Eggerthella lenta	×	○	○
Enterobacter cloacae	○	○	○
Enterococcus faecalis	○	○	○
Escherichia coli	○	○	○
Klebsiellla pneumoniae	○	○	○
Lactobacillus fermentum	○	○	○
Pseudomonas aeruginosa	○	○	○
Staphylococcus aureus	○	*S. simiae*	*S. haemolyticus*
Streptococcus pneumoniae	○	*S. pseudopneumoniae*	○

서 열 목 록

<110> Industry-Academic Cooperation Foundation, Yonsei University

Microbiotix Co., Ltd.

<120> Method for analyzing accuracy of next generation sequencing

<130> DPB174260

<160> 6

<170> KoPatentIn 3.0

<210> 1

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 1

tcgtcggcag cgtcagatgt gtataagaga cagagrgtty gatymtggct cag 53

<210> 2

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 2

gtctcgtggg ctcggagatg tgtataagag acaggctgcc tcccgtagga gt 52

<210> 3

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 3

tcgtcggcag cgtcagatgt gtataagaga cagcctacgg gnggcwgcag 50

<210> 4

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 4

gtctcgtggg ctcggagatg tgtataagag acaggactac hvgggtatct aatcc 55

<210> 5

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 5

tcgtcggcag cgtcagatgt gtataagaga cagaaactya aakraatwga cgg 53

<210> 6

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Primer of V1V2, V3V4, or V6V8

<400> 6

gtctcgtggg ctcggagatg tgtataagag acaggctgcc tcccgtagga gt 52