

2. Informatics - NCBI Blast search 결과 (HTML 파일 참조)

Reference:

Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000),
"A greedy algorithm for aligning DNA sequences", J Comput Biol 2000;
7(1-2):203-14.

Database: 16S Microbial Sequences
20,845 sequences; 30,310,001 total letters

Query= Contig_1

Length=1467

Sequences producing significant alignments:	Score (Bits)	E Value
NR_116070.1 Acinetobacter septicus strain AK001 16S ribosomal RNA...	2695	0.0
NR_116071.1 Acinetobacter septicus strain AK001 16S ribosomal RNA...	2673	0.0
NR_116069.1 Acinetobacter septicus strain AK001 16S ribosomal RNA...	2647	0.0
NR_025392.1 Acinetobacter ursingii strain LUH3792 16S ribosomal R...	2623	0.0
NR_102814.1 Acinetobacter oleivorans strain DRI 16S ribosomal RNA...	2488	0.0
NR_117621.1 Acinetobacter pittii DSM 21653 strain ATCC 19004 16S ...	2483	0.0
NR_042387.1 Acinetobacter calcoaceticus strain NCCB 22016 16S rib...	2479	0.0
NR_116774.1 Acinetobacter pittii DSM 21653 strain CIP 70.29 16S r...	2470	0.0
NR_134684.1 Acinetobacter seifertii strain LUH 1472 16S ribosomal...	2459	0.0
NR_116039.1 Acinetobacter oleivorans strain DRI 16S ribosomal RNA...	2447	0.0
NR_117630.1 Acinetobacter tandouii strain DSM 14970 16S ribosomal ...	2444	0.0
NR_117619.1 Acinetobacter calcoaceticus strain ATCC 23055 16S rib...	2444	0.0
NR_159919.1 Acinetobacter piscicola strain LW15 16S ribosomal RNA...	2440	0.0
NR_117629.1 Acinetobacter tjernbergiae strain DSM 14971 16S ribos...	2438	0.0
NR_117622.1 Acinetobacter haemolyticus strain ATCC 17906 16S ribo...	2438	0.0
NR_114958.1 Acinetobacter calcoaceticus strain ATCC 23055 16S rib...	2435	0.0
NR_117626.1 Acinetobacter guillouiae strain ATCC 11171 16S riboso...	2429	0.0
NR_117193.1 Acinetobacter tjernbergiae strain DSM 14971 16S ribos...	2429	0.0
NR_157606.1 Acinetobacter colistiniresistens strain NIPH 1859 16S...	2427	0.0
NR_152004.1 Acinetobacter lactucae strain NRRL B-41902 16S riboso...	2427	0.0
NR_117628.1 Acinetobacter bouvetii strain DSM 14964 16S ribosomal...	2427	0.0
NR_117620.1 Acinetobacter baumannii strain ATCC 19606 16S ribosom...	2427	0.0
NR_113343.1 Acinetobacter calcoaceticus strain JCM 6842 16S ribos...	2425	0.0

Informatics analysis 파일 중 assembly(=align), Blast search까지 결과 보내드립니다.
phylogenetic tree는 변수에 따라 너무 많이 달리 그려지고 분석되어서 제공되지 않습니다.

미생물동정 서비스 의뢰 시 보내드리는 데이터 예시파일



2. Informatics – 용어 설명

Blast search : Sequencing을 통해 얻은 염기서열을 NCBI (미국국립생물정보센터)의 Database내의 염기서열과 비교하여 유사성이 높은 종에 대한 결과 확인.

: 직접 Align 후 NCBI 홈페이지를 통해 Blast search를 수행할 수 있습니다.

: 주소 : <http://www.ncbi.nlm.gov/>

Blast Search 결과 확인

Sequences producing significant alignments:	Score (Bits)	E Value
gb GU294302.1 Stenotrophomonas maltophilia strain SJ-1 16S ribo...	2617	0.0
gb EU804401.1 Uncultured bacterium clone 6C232316 16S ribosomal...	2618	0.0
gb GU056295.1 Stenotrophomonas sp. enrichment culture clone ECC...	2618	0.0
gb EU805339.1 Uncultured bacterium clone 6C233343 16S ribosomal...	2618	0.0
gb EU708503.1 Uncultured bacterium clone ctg1_T0P01-17 16S ribo...	2618	0.0
gb EF572190.1 Uncultured bacterium clone S23_289 16S ribosomal ...	2618	0.0
gb DQ537219.1 Stenotrophomonas sp. YC-1 16S ribosomal RNA gene...	2618	0.0
gb DQ482654.1 Stenotrophomonas sp. Dsp-4 16S ribosomal RNA gene...	2618	0.0
emb AJ300772.1 Stenotrophomonas sp. LMG 19833 partial 16S rRNA ...	2618	0.0
gb JN995250.1 Stenotrophomonas sp. enrichment culture clone CW...	2517	0.0
gb GU557151.1 Stenotrophomonas sp. enrichment culture clone AOC...	2517	0.0
gb AF156709.1 AF156709 Stenotrophomonas sp. B0 16S ribosomal RNA...	2517	0.0
gb DQ298777.1 Bacterium 14A9 16S ribosomal RNA gene, partial se...	2517	0.0
gb GQ246699.1 Stenotrophomonas sp. M2T2B9 16S ribosomal RNA gen...	2515	0.0
gb JQ996678.1 Uncultured bacterium clone OTU-B3-2D 16S ribosoma...	2511	0.0
gb JQ996679.1 Uncultured bacterium clone OTU-B3-2 16S ribosoma...	2511	0.0

의뢰한 샘플과 database내 데이터를 비교 시 가장 유사한 종이지만 NGS처럼 분석 결과가 길지 않기 때문에 비슷한 것들이 많이 나열되어 있습니다. 아래에 있는 종일 수도 있습니다.

Score 값은 높을 수록 유사도가 높음
E-value (Expect-value) : 0에 가까울 수록 유사도가 높음



Score = 2063 bits (1117), Expect = 0.0
Identities = 1169/1195 (98%), Gaps = 15/1195 (1%)
Strand=Plus/Plus

NCBI blast search 단어와 수치들의 의미

<Score>

서열을 비교 검색할 때 특정 알고리즘(Smith-Waterman algorithm)을 사용하게 되는데 이때 각 서열마다 일정한 숫자를 부여하고 맨 나중에 이 숫자들을 취합하여 가장 높은 숫자를 가장 잘 맞는 비교로 순위를 매김.

<Expect>

현재 alignment와 동일한 결과가 나올 확률로서 0에 가까울수록 match가 잘 이루어짐을 나타냄.

<Identities>

서열의 유사도를 나타냄. 위의 그림에서 reference는 1195bp이며, blast search한 sequence는 1169bp로 유사도는 98%임. 100%에 가까울수록 match가 잘 이루어졌음을 나타냄

<Gaps>

서열을 비교하는 중간에 생겨난 gap. 예를 들어 query와 subject(NCBI reference)간의 mismatch가 일어난 부분을 이야기함, 아래와 같은 경우 gap이 2개 존재하므로 Gaps=2/10(20%)로 나타냄.

Ex)
query AGCTTATCGA
 ||| ||||
Subject AGGTTCTCGA

<Strand>

서열의 방향성을 보여줌. Reference는 항상 일정한 방향(Plus)로 이루어져 있지만 query sequence는 사용자마다 혹은 assembly 결과에 따라 Plus