MEGA basics

DATA alignment

- 1. Alignment -> Alignment Exploror/CLUSTAL -> create new alignment DNA form을 선택
- 2.1. FASTA form으로 정리되어 있는 시퀀스들을 정렬하려면 FASTA 파일을 열어 전체를 카피하여 위의 alignment 창에 복사하면 그대로 들어감.
- 2.2.1. 하나의 시퀀스를 이용하여 GenBank에 존재하는 비슷한 염기서열들을 검색하여 모든 뒤 정렬하려면 FASTA format의 하나의 시퀀스를 카피해서 입력 (꺽은쇄를 포함한 제목 도 같이; 여러개 시퀸스도 입력 가능)
- 2.2.2. 하나의 시퀀스를 하이라이트 시킨 후 왼쪽에서 다섯 번째 아이콘 "blast selected sequence"
- 2.2.3. blast 결과에서 원하는 시퀀스들을 체크박스에 체크 한 후 FASTA form text file로 전환 (체크박스에 체크 후 Send to text 를 클릭)
- 2.2.4 스크린 맨 위의 (+ Add to Alignment)를 누르면 Mega 프로그램에 자동적으로 채크 된 시퀀스가 입력됨
- 3. MEGA의 "Alignment Explorer" 창으로 다시 가서 Alignment -> Align by CLUSTALW

참고: MEGA에는 CLUSTALX의 구 버전인 CLUSTALW가 module로 들어있어서 MEGA 프로그램 내에서 alignment와 phylogenetic analysis 모두를 수행 할 수 있다.

- 4. Default option 값으로 alignment 시작 (gap opening penalty, gap extension penalty 등의 option을 조정할 수 있음)
- 5. 이때 눈으로 alignment 조정 가능
- 6. coding sequence (유전자 구간) 라면 왼쪽에서 4번째 icon을 이용하여 맞는 frame을
 확인 한 후 amino acid로 전환하여 amino acid data를 만들 수도 있음.
- 7. 원하는 형태의 데이터 (DNA 또는 protein)으로 전환 후 DATA -> Export Alignment -> MEGA format 을 선택하여 데이터 저장
 - cf) "input title"에는 분석의 제목을 임의로 넣으면 됨.
- 8. "Alignment Explorer"창을 닫고 MEGA 원래 창으로 돌아옴.

Distance calculation

- 1. File -> Open Data에서 저장한 데이터 파일을 갖고 옴.
- 2. View Sequence Data 창에서 왼쪽에서 네 번째 아이콘을 눌러 select/edit taxa group 창을 띄움.
- 3. 데이터 내에서 원하는 소그룹을 설정
- 4. Mega 창으로 돌아와서 Distances -> substitution model -> 원하는 모델 설정 후 OK 예) nucleotide -> Kimura 2-parameter
 - amino acid -> Pam matrix
- 5. 일대일 distance를 볼 때 Distance -> compute pairwise

전체 평균 distance를 볼 때 Distance -> compute overall mean

각 그룹 내의 평균 distance를 볼 때 Distance -> compute within group mean

그룹 간의 distance를 볼 때 Distance -> compute between groups means

그룹 간의 평균 distance를 볼 때 Distance -> compute net between groups means

Phylogeny reconstruction

1. File -> open data로 데이터를 load함

- 2. construct phylogeny -> 원하는 옵션 선택
- 3. 예로서 neighbor-joing 선택 -> 원하는 모델 선택 (ex. Kimura 2-parameter)
- 4. 그려진 tree를 root, tree order 등을 조정하여 최종적으로 image -> save as inhanced metafile (EMF)로 저장하면 저장된 파일을 파워포인트에 서 열 수 있음.
- 5. node들의 통계적인 지지도를 나타내기 위해서는 phylogeny -> bootstrap test of phylogeny에서 bootstrap replication 횟수를 지정해 주고 (통상 100번) 계산 하면 각 tree에서 각 node들의 bootstrap value들을 얻을 수 있다.