**EgPet NGS 알고리즘**

-------------------------------------EgPet NGS MIrROR 분석 파이프라인--------------------------------

1. **새로운 Sample sequencing data Input**
2. **Basecalling/Read correction**
3. **De-multiplexing/Adapter trimming**
4. **MIrROR process**
   1. Long read mapping
   2. Classification (Read count)
   3. Output = Mapping output file(paf 파일), Read count classification file(.txt, .std, mpa 파일)
5. **질환 별 MRS 지수 도출을 위한 taxonomy proportion 계산**

---------------------------------질환 별 MRS 계산 논리 및 프로세스-------------------------------

**참고:** <https://github.com/Gyungbu/EgPet>

**egpet\_percentile\_rank.py 파일**

1. **반려동물 종류 판별** 
   1. Proportion file의 이름을 기준으로 나눔 (ex, ‘PD’ 로 시작: dog, ‘PC’ 로 시작: cat)
   2. 분석할 샘플의 proportion file load
2. **파일 load**
   1. EGpet\_references.xlsx– 각 질환 별 미생물 리스트, 유익균, 유해균 리스트가 저장된 파일
   2. /input/db\_abundance\_{self.species}.csv – dog: 170마리, cat 59마리에 대한 Relative Abundance 값이 저장된 파일
   3. /input/egpet\_mrs\_db\_{self.species}.csv – dog: 170마리, cat 59마리에 대한 각 질환 별 MRS(Microbiome Risk Score), Dysbiosis, Healthy Distance, Diversity, Total Score 값이 저장된 파일
   4. /input/egpet\_percentile\_rank\_db\_{self.species}.csv – dog: 170마리, cat 59마리에 대한 각 질환 별 MRS(Microbiome Risk Score), Dysbiosis, Healthy Distance, Diversity, Total Score 값의 백분위 값이 저장된 파일
3. **MRS 계산**

: Microbiome Risk Score

: Relative abundance of microbiome associated with specific diseases

: The number of microbiome associated with a specific disease

: Health sign +1 for 유해/증가, -1 for 유익/감소

1. **Dysbiosis 계산**

: Among the microbiome in the dysbiosis\_microbiome\_{self.species}.xlsx file, the relative abundance of microbiome whose health sign is only harmful

: Among the microbiome in the dysbiosis\_microbiome\_{self.species}.xlsx file, the relative abundance of microbiome whose health sign is only beneficial

: The number of microbiome whose health sign is only harmful among microbiome in the dysbiosis\_microbiome\_{self.species}.xlsx

: The number of microbiome whose health sign is only beneficial among microbiome in the dysbiosis\_microbiome\_{self.species}.xlsx

1. **최종 Percentile Rank 계산**

앞서 구한 샘플의 MRS, Dysbiosis, Diversity를 기존 /input/egpet\_mrs\_db\_{self.species}.xlsx 파일을 기준으로 Percentile Rank 를 계산

scipy.stats.percentileofscore를 사용.

* 1. 백분위 값이 95 이상의 outliers 95로 처리, 5 이하의 outliers 5로 처리

1. **평가**
   1. 각 질환 별 , , , 값을 기준으로 (90,95] : 우수, (70,90] : 양호, (50,70] : 보통, (30,50] : 주의, [5,30] : 나쁨 등급으로 나눔.
   2. Diversity, Dysbiosis 값을 기준으로 Type E, B, I, D를 나눔

* Type E: Diversity 값의 백분위 (60/0.8)% 이상 & (-Healthy Distance) 값의 백분위와 (-Dysbiosis) 값의 백분위의 평균이 (60/1.1)% 이상인 경우
* Type B: Diversity 값의 백분위 (60/0.8)% 미만 & (-Healthy Distance) 값의 백분위와 (-Dysbiosis) 값의 백분위의 평균이 (60/1.1)% 이상인 경우
* Type I: Diversity 값의 백분위 (60/0.8)% 이상 & (-Healthy Distance) 값의 백분위와 (-Dysbiosis) 값의 백분위의 평균이 (60/1.1)% 미만인 경우
* Type D: Diversity 값의 백분위 (60/0.8)% 미만 & (-Healthy Distance) 값의 백분위와 (-Dysbiosis) 값의 백분위의 평균이 (60/1.1)% 미만인 경우

1. **유익균/유해균 abundance, number 계산**
   1. /input/dybiosis\_microbiome\_{self.species}.xlsx 파일의 유익균, 유해균 리스트를 기준으로 샘플의 유익균, 유해균 abundance[%] 값을 계산
   2. dog: 170마리, cat 59마리 데이터를 바탕으로 유익균, 유해균의 평균 abundance[%] 값을 계산

* num\_total\_species: proportion 파일의 observed 값
* num\_beneficial\_species: 유익균 리스트 중 MIrROR name의 species level 중 검출된 microbiome의 개수
* num\_harmful\_species: 유해균 리스트 중 MIrROR name의 species level 중 검출된 microbiome의 개수
* other\_abundance[%]:100-beneficial\_abundance[%]\_evaluation-harmful\_abundance[%]\_evaluation
* num\_other\_species: 장내에\_있던\_세균종수 - 유익균의\_세균종수 - 유해균의\_세균종수