02_OPTICS_predict_lmax

September 2, 2025

[1]:
```
import sys
from pathlib import Path

#Define paths for current project
# --- Centralized paths ---
ROOT = Path("..")
DATA = ROOT / "data"
LOGS = ROOT / "logs"
SCRIPTS = ROOT / "scripts"
RESULTS = ROOT / "results"
ALIGN_DIR = RESULTS / "align"
TREE_DIR = RESULTS / "trees"
FIGURES = RESULTS / "figures"


# Set the path to your local optics codebase
optics_path = str(Path.home() / "labdata/users/Oakley/GitHub/optics")
if optics_path not in sys.path:
    sys.path.append(optics_path)

from optics_predictions import run_optics_predictions

# Input FASTA: deduplicated, combined rhodopsin sequences
fasta_file = DATA / "pumphits_culled.fasta"
results_dir = "../results/optics"

# Run OPTICS predictions with options matching the CLI help
optics_df, optics_pred_file = run_optics_predictions(
    input_sequence=fasta_file,
    pred_dir=results_dir,                          # --output_dir
    output="optics_predictions",                   # --prediction_prefix
    model="type-one",                              # --model
    encoding_method="aa_prop",                     # --encoding
    blastp=False,                                  # --blastp
    iden_report="blastp_report.txt",               # --blastp_report
    refseq="bovine",                               # --refseq
    bootstrap=False                                # --bootstrap
```

```
)

print(f"OPTICS predictions saved to: {optics_pred_file}")
optics_df.head()
```

```
Processing Sequences: 100%|              |
90/90 [00:27<00:00,  3.27seqs/s]
```

```
Error: Cached prediction file can't be saved…
```

```
YCyR2hit__Nodosilinea_nodulosa__UniRef90_A0AAJ6N6B2     544.6   -       256

YCyR2hit__Pseudanabaena_sp_FACHB_2040__UniRef90_A0A926ZYX4      532.5   -
235

YCyR2hit__Leptolyngbya_sp_FACHB_261__UniRef90_A0A926UF78        528.7   -
229

YCyR2hit__Hassalia_byssoidea_VB512170__UniRef90_A0A846HFU5      537.3   -
233

YCyR2hit__Halorientalis__UniRef90_A0A1G7PUC9    552.4   -       245

YCyR2hit__UnknownTaxon__UniRef90_P29563 546.8   -       259

YCyR2hit__Leptolyngbya_ohadii__UniRef90_UPI000B5A185C   568.0   -       236

YCyR2hit__Aliterella_atlantica_CENA595__UniRef90_A0A0D8ZW42     542.8   -
236

YCyR2hit__Cyanobacteriota__UniRef90_A0A926PVG5  554.1   -       232

YCyR2hit__Microcoleus_sp_FACHB_1515__UniRef90_A0A926U182        534.7   -
233

YCyR2hit__Halorientalis_marina__UniRef90_UPI001FF6C9D7  539.8   -       243

YCyR2hit__unclassified_Leptolyngbya__UniRef90_UPI0016884D19     556.9   -
235

YCyR2hit__Leptolyngbya_sp_NIES_2104__UniRef90_A0A0P4V0S4        557.0   -
236

YCyR2hit__unclassified_Nostoc__UniRef90_A0A252DUV9      538.5   -       234

YCyR2hit__Natronomonas_moolapensis__UniRef90_M1Y5I7     546.8   -       258

BRhit__Halobacterium__UniRef90_P02945   561.7   -       262
```

BRhit__Halobacterium__UniRef90_UPI001F011AB0 562.5 - 262

BRhit__Haloplanus__UniRef90_A0A6B9F3P5 558.1 - 256

BRhit__Haloplanus_natans__UniRef90_UPI000A5CE34D 562.1 - 256

BRhit__Halobaculum_rubrum__UniRef90_UPI001CA389E1 556.6 - 248

BRhit__Halohasta_salina__UniRef90_UPI002110EEC8 551.6 - 263

BRhit__Halobacteriales_archaeon_QS_8_69_26__UniRef90_A0A2R6JS63 548.4 - 265

BRhit__Uncultured_archaeon__UniRef90_I1X958 545.3 - 259

BRhit__Halorientalis_regularis__UniRef90_A0A1G7RA78 554.0 - 258

BRhit__unclassified_Halobacteriales__UniRef90_A0A2R6LDY7 545.8 - 264

BRhit__uncultured_archaeon_A07HR67__UniRef90_V4ZQH6 553.3 - 260

BRhit__Halobaculum_sp_MBLA0143__UniRef90_UPI00352426DC 556.7 - 260

BRhit__Halorubrum__UniRef90_A0A4U7F7H9 553.2 - 258

BRhit__Halobacteria__UniRef90_P69051 558.0 - 260

GCyR2hit__Nostocales__UniRef90_A0A252D874 543.6 - 247

GCyR2hit__Tolypothrix_bouteillei_VB521301__UniRef90_A0A8S9SVI8 543.3 - 247

GCyR2hit__Nostoc_sp__UniRef90_UPI002FFB8FFD 542.3 - 248

GCyR2hit__Tolypothrix_sp_NIES_4075__UniRef90_A0A218QMM7 539.6 - 252

GCyR2hit__Pleurocapsa_sp_FMAR1__UniRef90_UPI0029C671DC 541.8 - 246

GCyR2hit__Nostocaceae_cyanobacterium__UniRef90_A0A838VMV4 516.6 - 244

GCyR2hit__Phormidesmis_priestleyi__UniRef90_UPI0009442622 542.9 - 248

GCyR2hit__Pseudanabaenaceae_cyanobacterium_LEGE_13415__UniRef90_A0A928V9X3 542.6 - 248

GCyR2hit__Cyanobacteria_bacterium_QH_9_48_43__UniRef90_A0A2T2RN81 542.0 - 234

GCyR2hit__Halegenticoccus_tardaugens__UniRef90_UPI00100C2973 542.0 - 241

GCyR2hit__Haloferax_mucosum_ATCC_BAA_1512__UniRef90_M0IDG5 547.7 - 243

GCyR2hit__Leptolyngbya_sp_FACHB_261__UniRef90_A0A926UF78 528.7 - 229

GCyR2hit__Halobiforma__UniRef90_A0A1P8LVS0 545.0 - 240

GCyR2hit__Halogranum_amylolyticum__UniRef90_A0A1H8RSA7 546.2 - 241

GCyR2hit__Halobacteria__UniRef90_O93740 547.0 - 250

GRhit__Gloeobacter_violaceus__UniRef90_Q7NP59 539.0 - 298

GRhit__Cyanobacteria_bacterium_RM1_2_2__UniRef90_A0A969T0G4 538.5 - 299

GRhit__Cyanophyceae__UniRef90_A0A2W7ARY7 538.9 - 297

GRhit__unclassified_Leptolyngbyaceae__UniRef90_A0A969FEC7 538.5 - 297

GRhit__Leptolyngbya_sp_ES_bin_22__UniRef90_A0A925M2S9 540.3 - 297

GRhit__Phormidesmis_priestleyi__UniRef90_UPI000A475A3D 539.6 - 298

GRhit__Chamaesiphon_sp__UniRef90_UPI003592F9D2 538.1 - 304

GRhit__Leptolyngbya_sp_Heron_Island_J__UniRef90_U9W0I1 541.7 - 282

GRhit__Cyanophyceae__UniRef90_A0A8J7JTD9 544.0 - 269

GRhit__Halomicronema_sp_CCY15110__UniRef90_UPI0021030F7B 543.2 - 281

GRhit__Cyanophyceae__UniRef90_A0A4Q7E4T5 543.4 - 281

GRhit__Leptolyngbyaceae_cyanobacterium_JSC_12__UniRef90_K8GTY7 538.7 - 262

GRhit__Halothece_sp__UniRef90_K9YEI2 543.6 - 281

GRhit__Leptolyngbya_sp_LCM1_Bin17__UniRef90_A0A6H2NHV9 540.0 - 276

GRhit__Symploca_sp_SIO2G7__UniRef90_A0A845YHD9 538.7 - 265

GRhit__Deinococcus__UniRef90_A0A172TD44 545.0 - 285

GRhit__Deinococcus_ruber__UniRef90_A0A918CFF3 543.7 - 293

GRhit__Deinococcus__UniRef90_A0A917PC33 541.0 - 295

GRhit__Deinococcus_sp__UniRef90_UPI0025F8D0E5 534.1 - 259

GRhit__Allomeiothermus_silvanus__UniRef90_UPI0023F38943 541.4 - 258

GRhit__Bdellovibrio_sp__UniRef90_A0A924B300 544.2 - 253

GRhit__Trueperaceae_bacterium__UniRef90_A0A5Q4G3V2 543.2 - 264

GRhit__Myxococcales_bacterium__UniRef90_A0A2E7ZR97 537.5 - 261

GRhit__Thermaceae_bacterium__UniRef90_A0A931EK09 534.2 - 262

GRhit__Oligoflexia_bacterium__UniRef90_A0A923U6H7 550.1 - 255

GRhit__Deinococcus_sp__UniRef90_UPI002869BBA0 542.5 - 299

GRhit__Bacteroidota_bacterium__UniRef90_A0A3M2FMH9 542.2 - 260

GRhit__Rhodothermales_bacterium__UniRef90_A0A9D8YSH6 538.4 - 261

GRhit__Pseudobdellovibrionaceae_bacterium__UniRef90_A0A924ZXQ0 548.8 - 253

GRhit__Deinococcus_koreensis__UniRef90_A0A2K3UTP4 537.1 - 285

PRhit__Bacteria__UniRef90_Q6J4G7 520.4 - 250

PRhit__uncultured_bacterium__UniRef90_Q84C21 517.5 - 258

PRhit__SAR86_cluster_bacterium__UniRef90_A0A937M2P1 519.3 - 250

PRhit__environmental_samples__UniRef90_Q6PL05 520.8 - 251

PRhit__uncultured_bacterium__UniRef90_A0A1L2YW20 500.0 - 251

PRhit__Bacteria__UniRef90_Q9AFF7 495.9 - 251

```
PRhit__Bacteria__UniRef90_A0A1L2YW54    495.4   -       249

PRhit__Gammaproteobacteria__UniRef90_A0A368C6X1 506.3   -       250

PRhit__Gammaproteobacteria_bacterium_TMED104__UniRef90_A0A3S5JK93       526.2
-       247

PRhit__Bacteria__UniRef90_J4KSW5        521.3   -       251

PRhit__Bacteria__UniRef90_A0A2E1JM07    504.6   -       250

PRhit__environmental_samples__UniRef90_A0A1L2YWB0       520.2   -       250

PRhit__Pseudomonadota_bacterium__UniRef90_A0A944UJK0    520.1   -       249

PRhit__Pseudomonadota__UniRef90_A0A0R2UEA1      525.2   -       248

PRhit__Gammaproteobacteria_bacterium__UniRef90_A0A2E8KSB6       506.1   -
251

BRmutant_WP_136361479.1_D85A    610.2   -       263


Model Used:     type-one
Encoding Method:        aa_prop

Predictions Complete!
OPTICS predictions saved to: ../results/optics/optics_on_optics_predictions_2025
-09-01_21-43-10/optics_predictions_predictions.tsv
```

[1]:

	Names	Single_Prediction	%Identity_Nearest_VPOD_Sequence	Sequence_Length	Lmax_Hex_Color
0	YCyR2hit__Nodosilinea_nodulosa__UniRef90_A0AAJ…	544.6	-	256	#91ff00
1	YCyR2hit__Pseudanabaena_sp_FACHB_2040__UniRef9…	532.5	-	235	#67ff00
2	YCyR2hit__Leptolyngbya_sp_FACHB_261__UniRef90_…	528.7	-	229	#59ff00
3	YCyR2hit__Hassalia_byssoidea_VB512170__UniRef9…	537.3	-	233	#78ff00
4	YCyR2hit__Halorientalis__UniRef90_A0A1G7PUC9	552.4	-	245	#abff00

[2]:
```
#Plot histograms for sequence length and OPTICS predictions of lambda-max

%matplotlib inline
```

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the OPTICS results TSV
optics_tsv = RESULTS / "optics/optics_on_optics_predictions_2025-09-01_21-43-10/
 ↪optics_predictions_predictions.tsv"
df = pd.read_csv(optics_tsv, sep="\t")

# Plot histograms of lambda-max predictions and sequence length
fig, axes = plt.subplots(1, 2, figsize=(14, 5))

# Histogram for lambda-max predictions
df["Single_Prediction"].hist(bins=30, color="skyblue", edgecolor="black",␣
 ↪ax=axes[0])
axes[0].set_xlabel("Predicted  max (nm)")
axes[0].set_ylabel("Count")
axes[0].set_title("Distribution of OPTICS  max Predictions")

# Histogram for sequence length
df["Sequence_Length"].hist(bins=30, color="salmon", edgecolor="black",␣
 ↪ax=axes[1])
axes[1].set_xlabel("Sequence Length (aa)")
axes[1].set_ylabel("Count")
axes[1].set_title("Distribution of Sequence Lengths")

plt.tight_layout()
plt.show()
```