Exploring BLAST and PSI-BLAST to Detect Distant Homologs

1) In the initial stage, we obtained data related to the kinase sequence named "Myosin Light Chain Kinase Family Member 4 (MYLK4)" from UniProt (Figure 1). UniProt ID: Q86YV6 (https://www.uniprot.org/uniprotkb/Q86YV6/entry)

MAST4	MAST4	MAST4	Microtubule-associated serine/threonine-protein kinase 4	AGC	MAST		O15021
CaMK2d	CaMK2d	CAMK2D	Calcium/calmodulin-dependent protein kinase type II subunit delta	CAMK	CAMK2		Q13557
BIKE	BIKE	BMP2K	BMP-2-inducible protein kinase	Other	NAK		Q9NSY1
TSSK1	TSSK1	TSSK1B	Testis-specific serine/threonine-protein kinase 1	CAMK	TSSK		Q9BXA7
SgK424	SgK424	TEX14	Inactive serine/threonine-protein kinase TEX14	Other	NKF5		Q8IWB6
SgK085	SgK085	MYLK4	Myosin light chain kinase family member 4	CAMK	MLCK		Q86YV6
ADCK2	ADCK2	ADCK2	Uncharacterized aarF domain-containing protein kinase 2	Atypical	ABC1	ABC1-C	Q7Z695
DAPK3	DAPK3	DAPK3	Death-associated protein kinase 3	CAMK	DAPK		O43293

Figure 1. We learned to Myosin Light Chain Kinase Family Member 4 (MYLK4)'s Uniprot ID.

We displayed the information of MYLK4 sequenceF (Figure2).

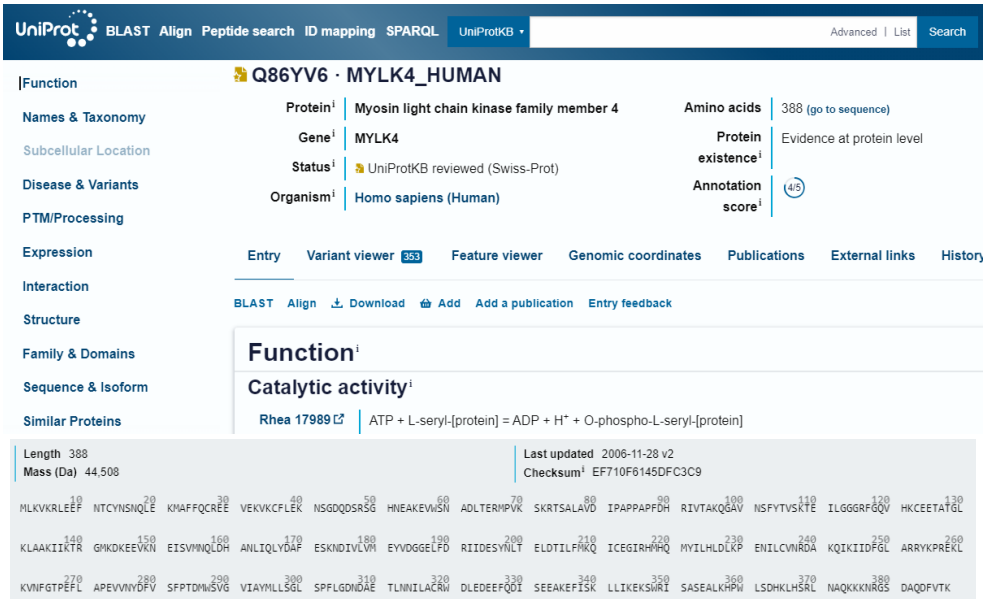


Figure 2. Q86YV6 displayed on UniProt Database.

We displayed the MYLK4 on KinMap (Figure2).

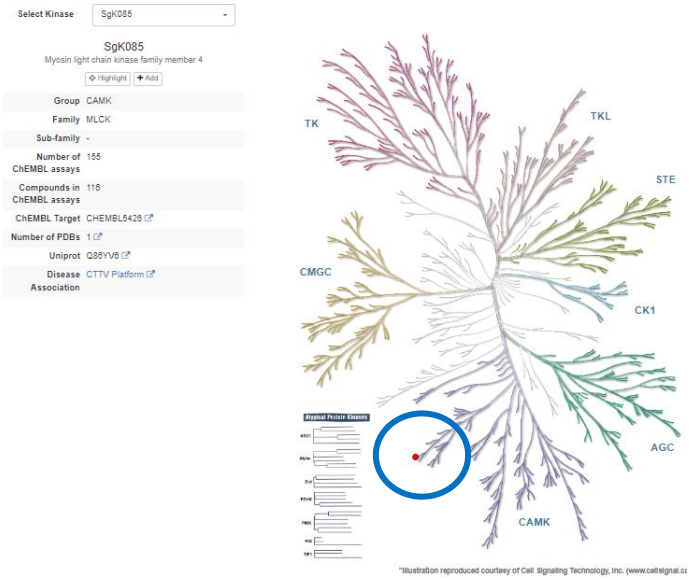


Figure 3. MYLK4 view on KinMap.

2) Using the obtained sequence information as input and selecting the relevant options, we initiated a PSI-BLAST search (Figure 4). The options that need to be selected:

Database: UniProtKB/Swiss-Prot
Organism: Homo sapiens
Max target sequences: 20
Matrix: PAM250

Parameters other than these options are left as default values.

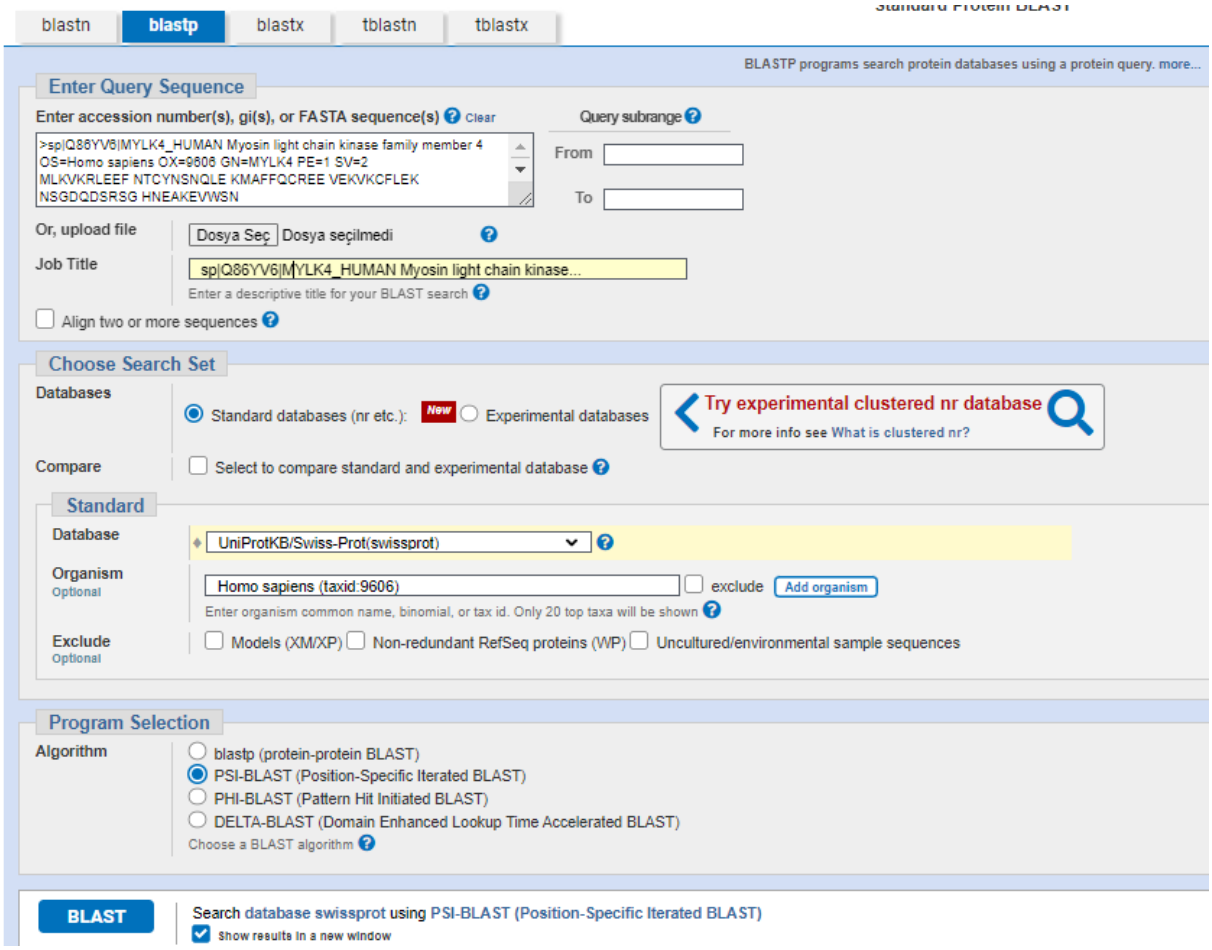


Figure 4. We start the PSI-BLAST search as the options.

3) In the previous step, the 'max target sequence' option was set to 10 (the option for 20 was not available for the first iteration). We downloaded the output table in text format for the first iteration. From this table, we stored the UniProt IDs provided for each kinase in a table. Subsequently, using these IDs, we performed a list-style search on UniProt and reached the corresponding information for each kinase.

Subsequently, we added the 10 kinases obtained in the first stage to KinMap in list format and visualized their positions on the phylogenetic tree.

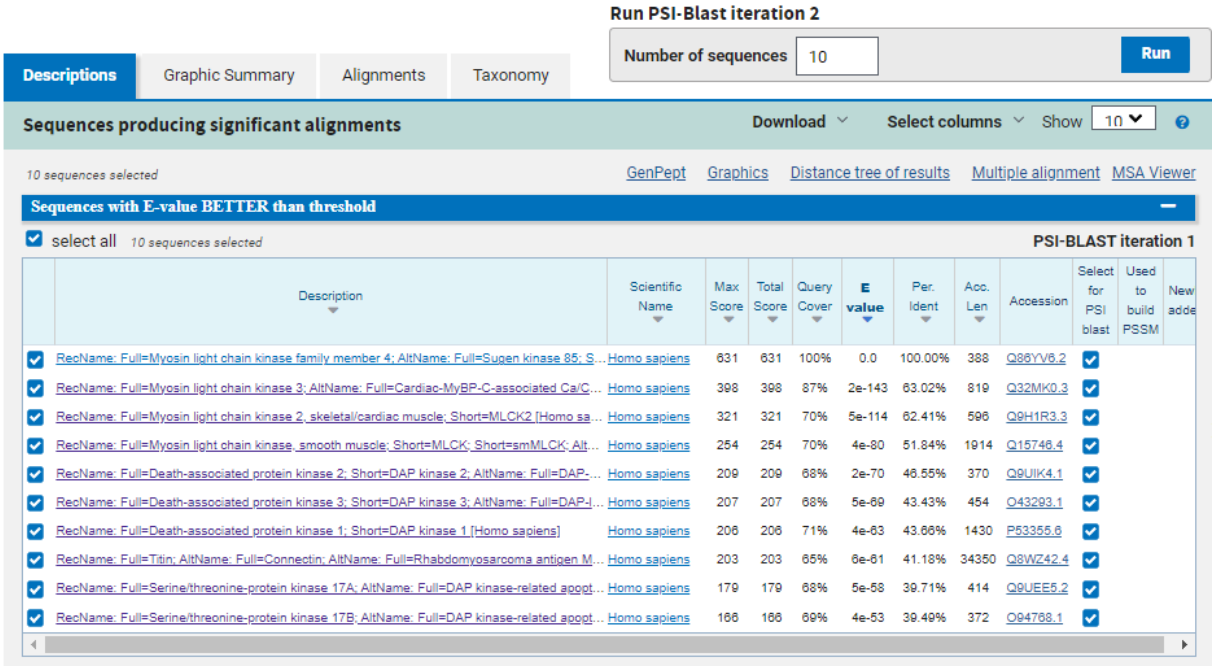
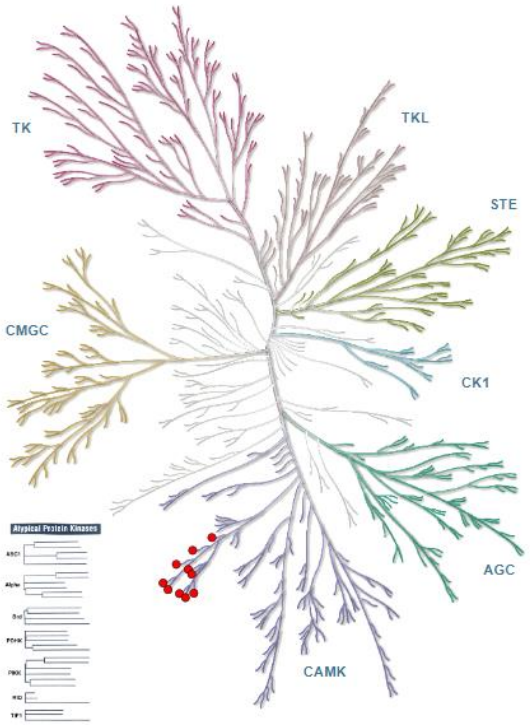


Figure 5. Result of first iteration.

Iteration 1: Max target sequnce = 10

TTN	Titin
STK17B	Serine/Threonine Kinase 17B
STK17A	Serine/Threonine Kinase 17A
MYLK	Myosin Light Chain Kinase
MYLK4	Myosin Light Chain Kinase 4
MYLK3	Myosin Light Chain Kinase 3
MYLK2	Myosin Light Chain Kinase 2
DAPK3	Death-Associated Protein Kinase 3
DAPK2	Death-Associated Protein Kinase 2
DAPK1	Death-Associated Protein Kinase 1



4) In this stage, we set the maximum target sequence for the second iteration to 20 and conducted a new search. We included additional hits in the result table and downloaded the UniProt IDs for the first iteration in text format, incorporating them into a table. Subsequently, by searching the UniProt IDs in list format, we reached the relevant kinases. The kinases included in the result table for Iteration 2 are highlighted in yellow (Figure 6).

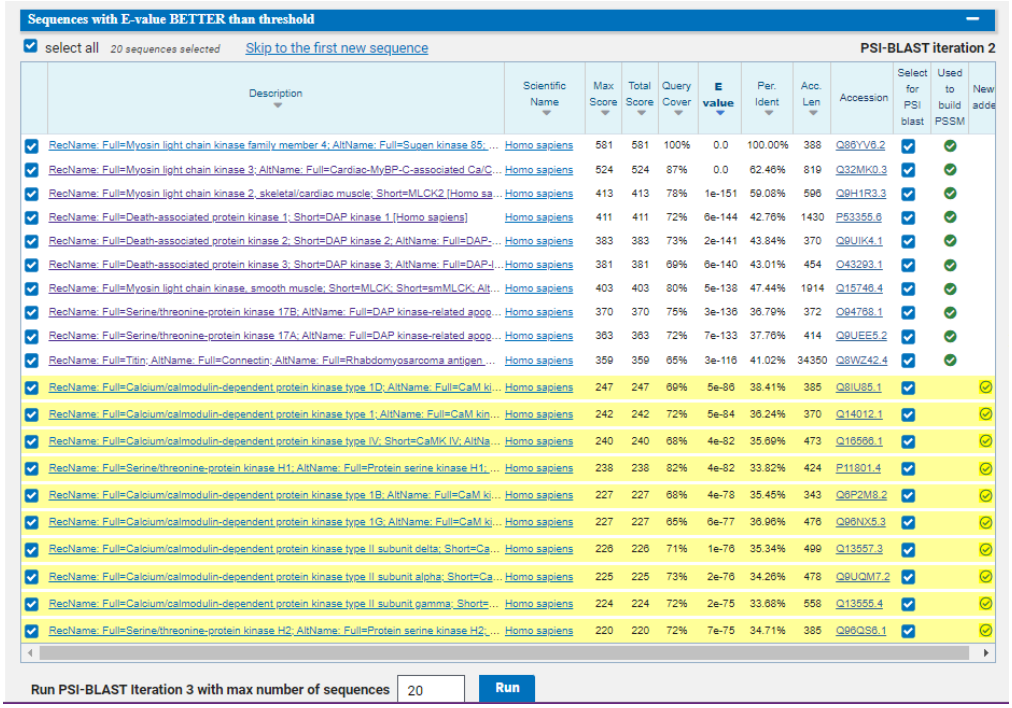
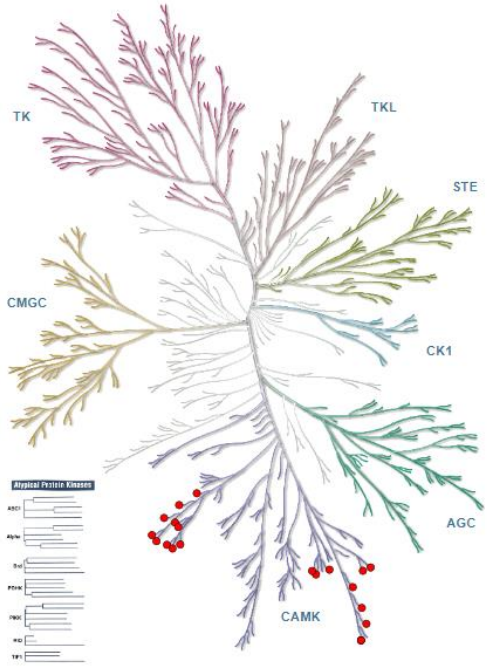


Figure 6. Result of second iteration and added 10 kinase.

Iteration 2: Max target sequnce = 20

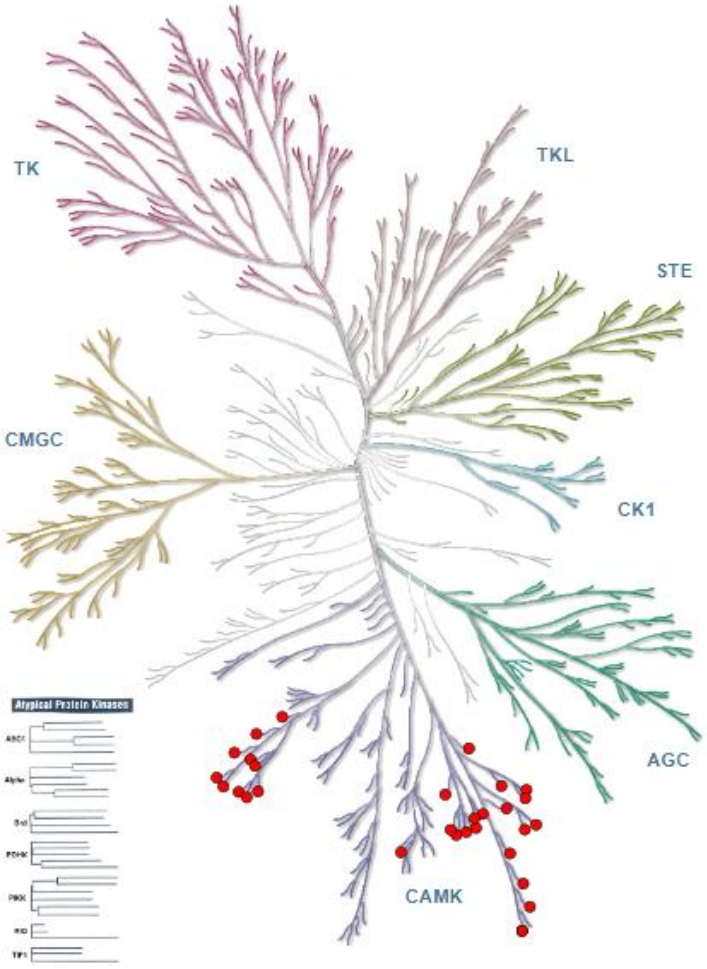
CAMK1D	Calcium/Calmodulin-Dependent Protein Kinase 1D
CAMK1D	Calcium/Calmodulin-Dependent Protein Kinase 1D
CAMK4	Calcium/Calmodulin-Dependent Protein Kinase 4
PSKH1	Protein Serine Kinase H1
PNCK	Pregnancy Upregulated Nonubiquitous CaM Kinase
CAMK1G	Calcium/Calmodulin-Dependent Protein Kinase 1G
CAMK2D	Calcium/Calmodulin-Dependent Protein Kinase 2D
CAMK2A	Calcium/Calmodulin-Dependent Protein Kinase 2A
CAMK2G	Calcium/Calmodulin-Dependent Protein Kinase 2G
PSKH2	Protein Serine Kinase H2



5) In all subsequent stages, the increment steps from the previous two iterations have been applied. The max target sequence has been increased by 10 at each step. Therefore, for the 3rd iteration, our max target sequence count is 30.

Iteration 3: Max target sequnce = 30

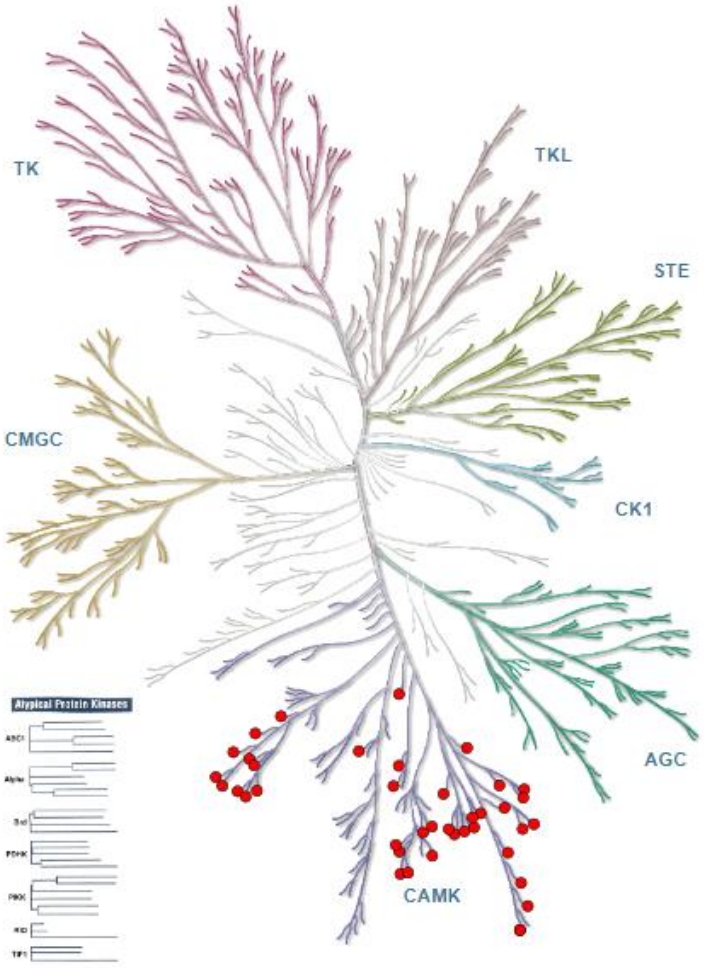
CAMK2B	Calcium/Calmodulin-Dependent Protein Kinase 2B
DCLK1	Doublecortin Like Kinase 1
DCLK2	Doublecortin Like Kinase 2
DCLK3	Doublecortin Like Kinase 3
CAMKV	Calcium/Calmodulin-Dependent Protein Kinase-Like Vesicle-Associated
PHKG1	Phosphorylase Kinase Catalytic Subunit Gamma 1
PHKG2	Phosphorylase Kinase Catalytic Subunit Gamma 2
CHEK2	Checkpoint Kinase 2
CASK	Calcium/Calmodulin-Dependent Serine Protein Kinase
RPS6KA2	Ribosomal Protein S6 Kinase A2



6) PSI-BLAST iterations were continued until kinases from at least 3 groups were included in the results table. The max target sequence was increased by 10 in each iteration. Below are the kinases included in the results table for the 4th iteration and subsequent iterations, along with their positions on the phylogenetic tree.

Iteration 4: Max target sequnce = 40

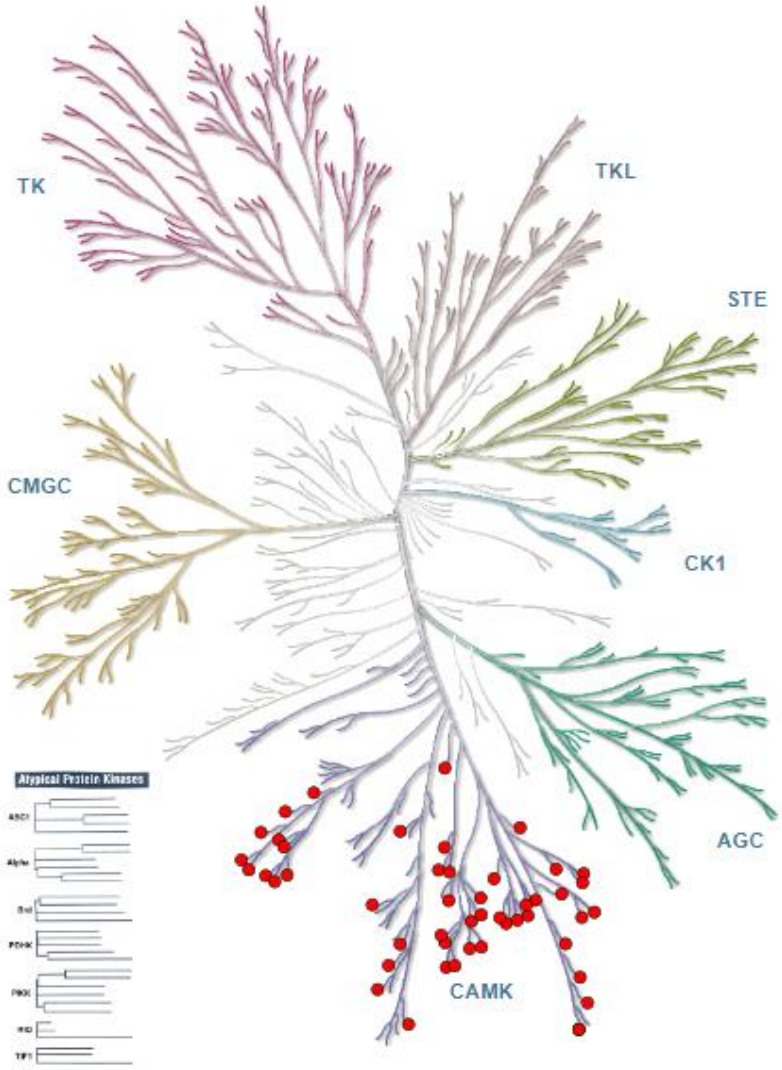
RPS6KA3	Ribosomal Protein S6 Kinase A3
RPS6KA1	Ribosomal Protein S6 Kinase A1
RPS6KA6	Ribosomal Protein S6 Kinase A6
PRKD2	Protein Kinase D2
RPS6KA5	Ribosomal Protein S6 Kinase A5
MAPKAPK3	Mitogen-Activated Protein Kinase-Activated Protein Kinase 3
PRKD1	Protein Kinase D1
STK33	Serine/Threonine Kinase 33
MAPKAPK2	Mitogen-Activated Protein Kinase-Activated Protein Kinase 2
SNRK	Sucrose Non-Fermenting Related Kinase



Iteration 5: Max target sequnce = 50

PRKD3	Protein Kinase D3
RPS6KA4	Ribosomal Protein S6 Kinase A4
MAPKAPK5	Mitogen-Activated Protein Kinase-Activated Protein Kinase 5
NUAK1	NUAK Family SNF1-Like Kinase 1
SIK1	Salt-Inducible Kinase 1

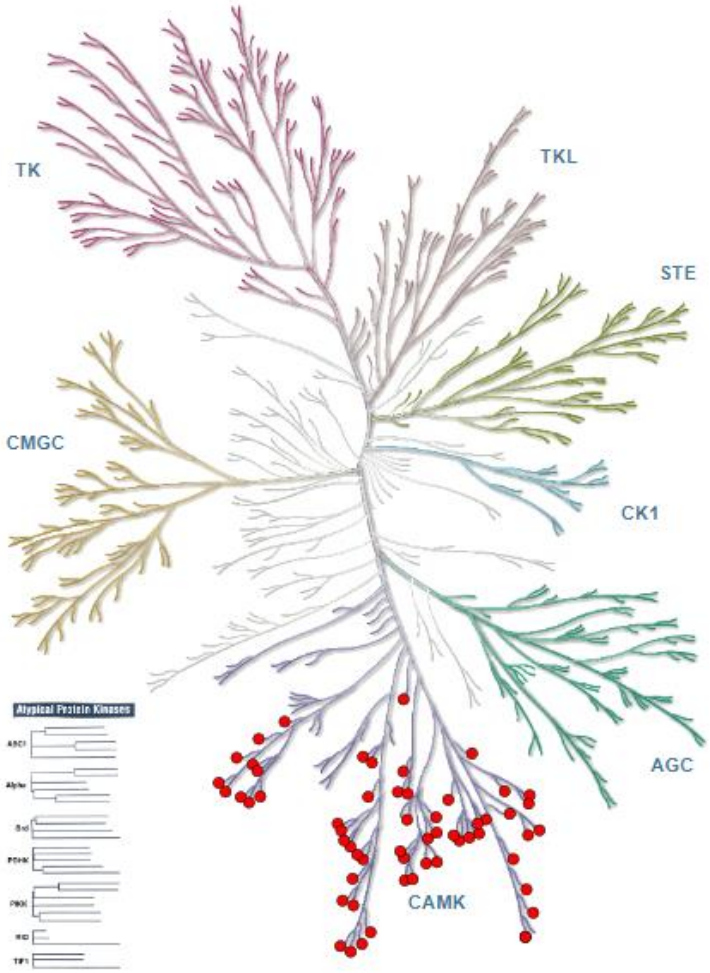
SIK1B	Salt-Inducible Kinase 1B
MKNK2	MAP Kinase-Interacting Serine/Threonine Kinase 2
SIK3	Salt-Inducible Kinase 3
MARK4	Microtubule Affinity Regulating Kinase 4
PRKAA2	Protein Kinase AMP-Activated Catalytic Subunit Alpha 2



Iteration 6: Max target sequnce = 60

MARK3	Microtubule Affinity Regulating Kinase 3
MARK2	Microtubule Affinity Regulating Kinase 2
SIK2	Salt-Inducible Kinase 2
PRKAA1	Protein Kinase AMP-Activated Catalytic Subunit Alpha 1

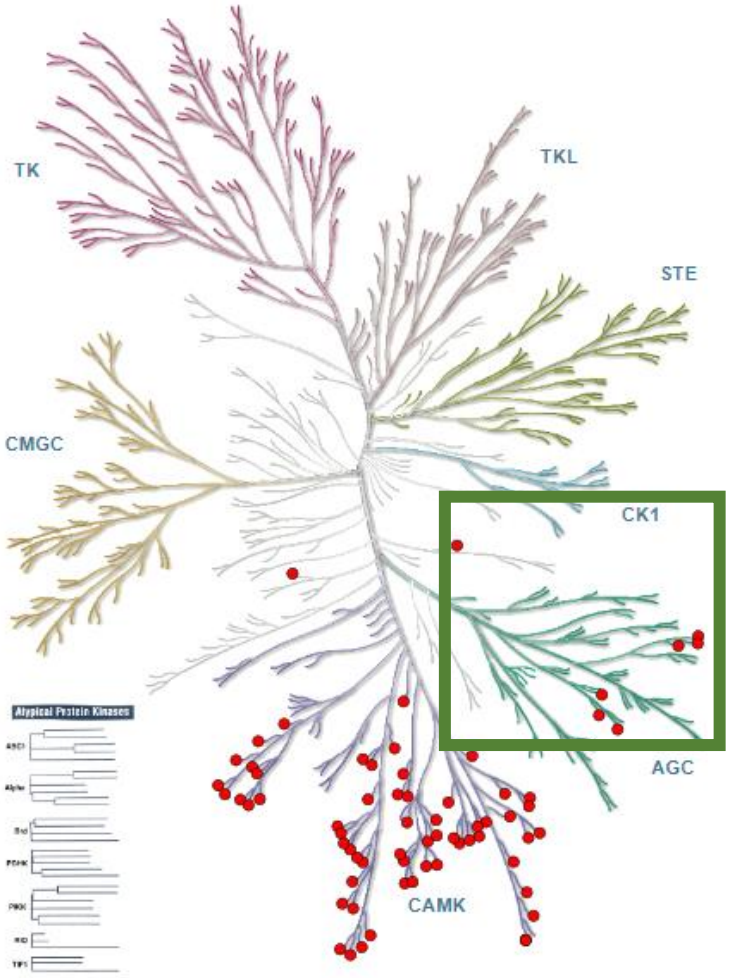
NUAK2	NUAK Family SNF1-Like Kinase 2
MARK1	Microtubule Affinity Regulating Kinase 1
BRSK1	Breast Tumor Kinase 1
MELK	Maternal Embryonic Leucine Zipper Kinase
BRSK2	Breast Tumor Kinase 2
NIM1K	Never in Mitosis Gene A-Related Kinase



Iteration 7: Max target sequnce = 70

In the 7th iteration, kinases from the AGC kinase group were observed for the first time outside the CAMK kinase group. The AGC group represents the protein kinase A, G, and C subgroups**.**

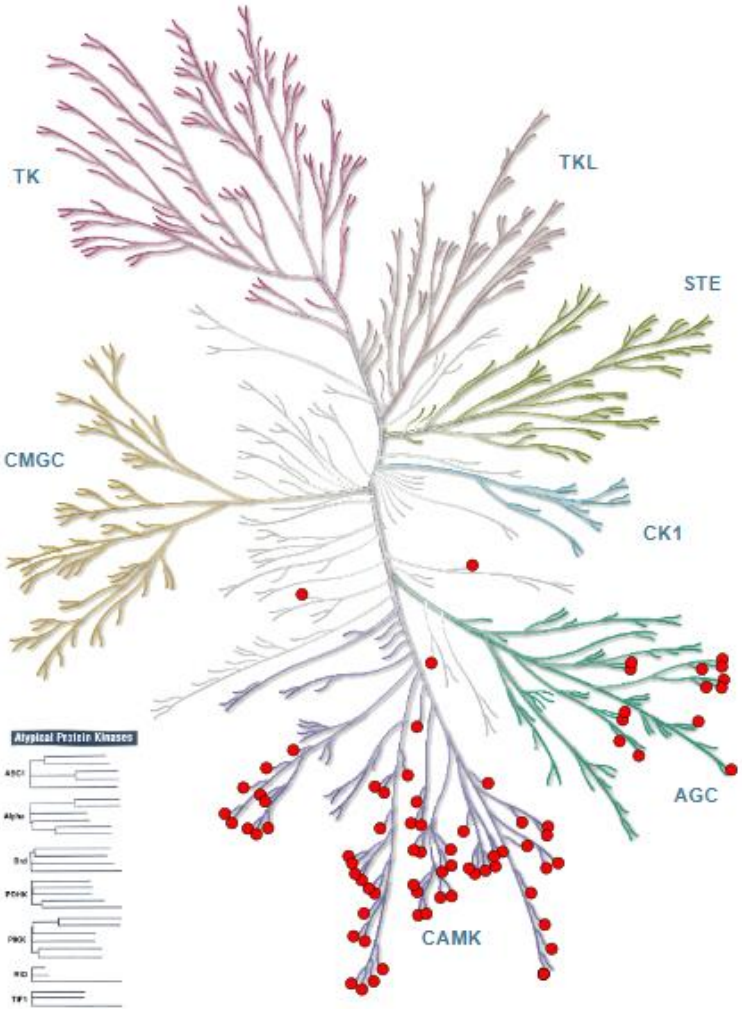
HUNK	Hormonally Upregulated Neu-associated Kinase
PRKACB	Protein Kinase CAMP-Activated Catalytic Subunit Beta
ORKACA	Ovarian Cancer Kinase A
AKT3	AKT Serine/Threonine Kinase 3
PLK4	Polo-Like Kinase 4
PRKACG	Protein Kinase CAMP-Activated Catalytic Subunit Gamma
AKT2	AKT Serine/Threonine Kinase 2
AKT1	AKT Serine/Threonine Kinase 1
PRKX	Protein Kinase X-Linked
ULK3	Unc-51 Like Autophagy Activating Kinase 3



Iteration 8: Max target sequnce = 80

PRKY	Protein Kinase Y-Linked
MKNK1	MAP Kinase-Interacting Serine/Threonine Kinase 1
SGK2	Serum/Glucocorticoid Regulated Kinase 2

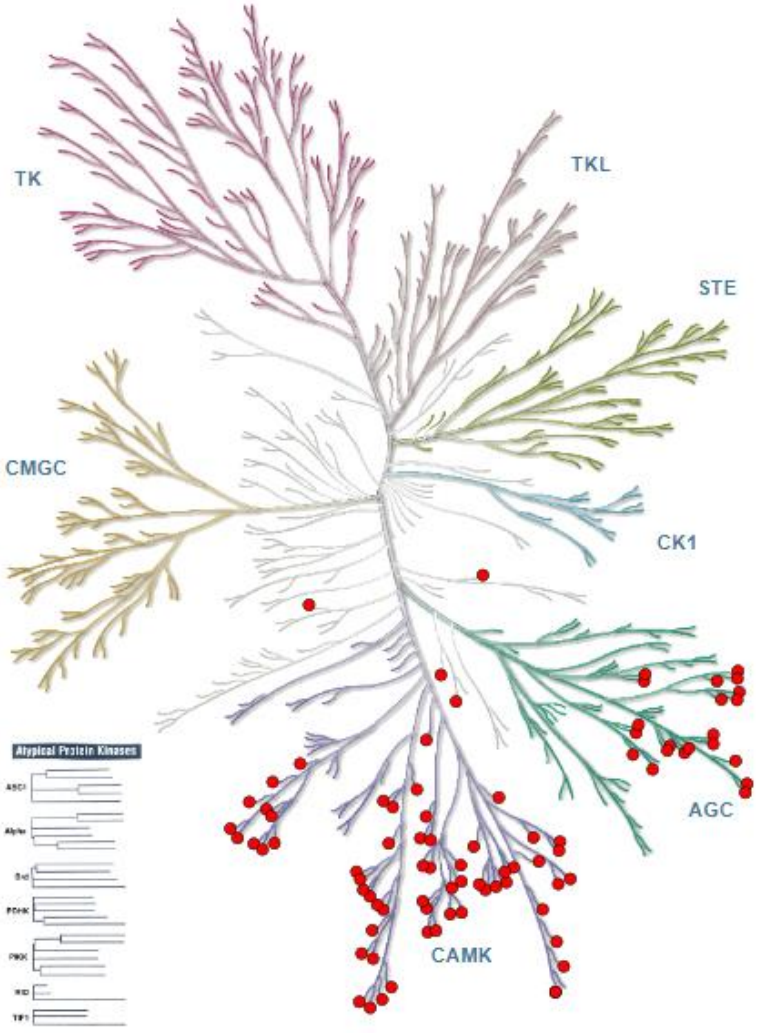
SGK1	Serum/Glucocorticoid Regulated Kinase 1
RPS6KB2	Ribosomal Protein S6 Kinase B2
RPS6KB1	Ribosomal Protein S6 Kinase B1
PRKCH	Protein Kinase C Epsilon
AURKA	Aurora Kinase A
SGK3	Serum/Glucocorticoid Regulated Kinase 3
PRKCA	Protein Kinase C Alpha



Iteration 9: Max target sequnce = 90

PRKCE	Protein Kinase C Epsilon
PRKCB	Protein Kinase C Beta

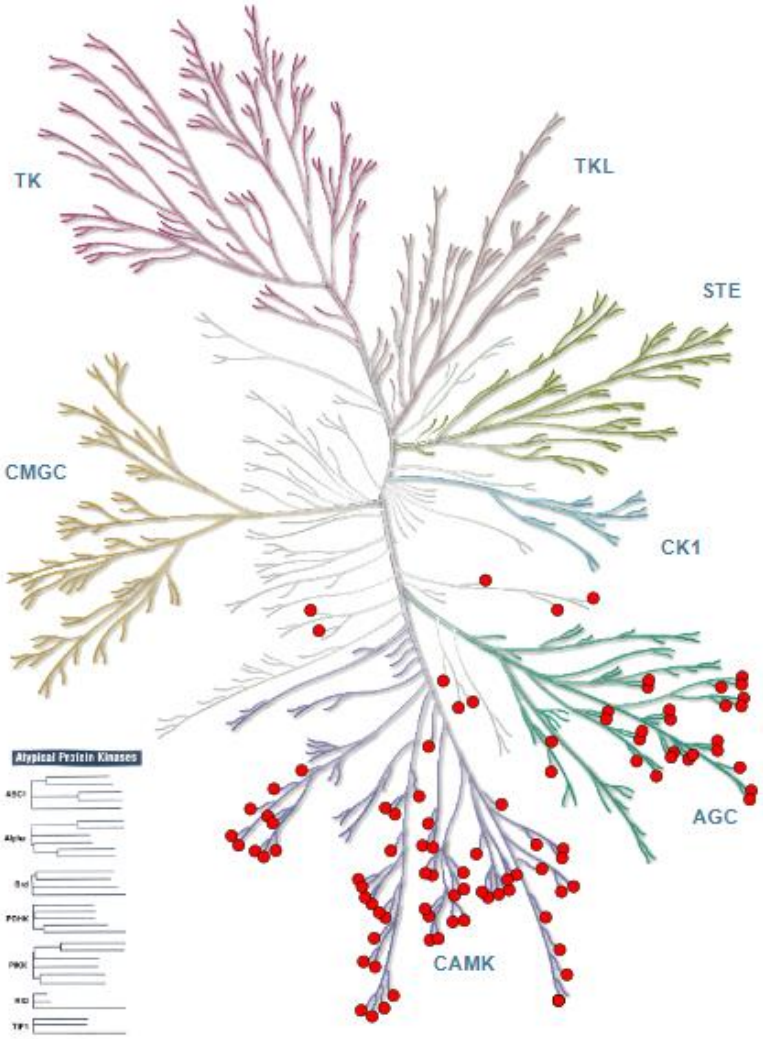
PRKCG	Protein Kinase C Gamma
AURKC	Aurora Kinase C
PRKCQ	Protein Kinase C Theta
PRKCI	Protein Kinase C Iota
PRKCD	Protein Kinase C Delta
PDPK2P	3-phosphoinositide dependent protein kinase-2 pseudogene
PRKCZ	Protein Kinase C Zeta
PDKP1	3-phosphoinositide dependent protein kinase-1 pseudogene



Iteration 10: Max target sequnce = 100

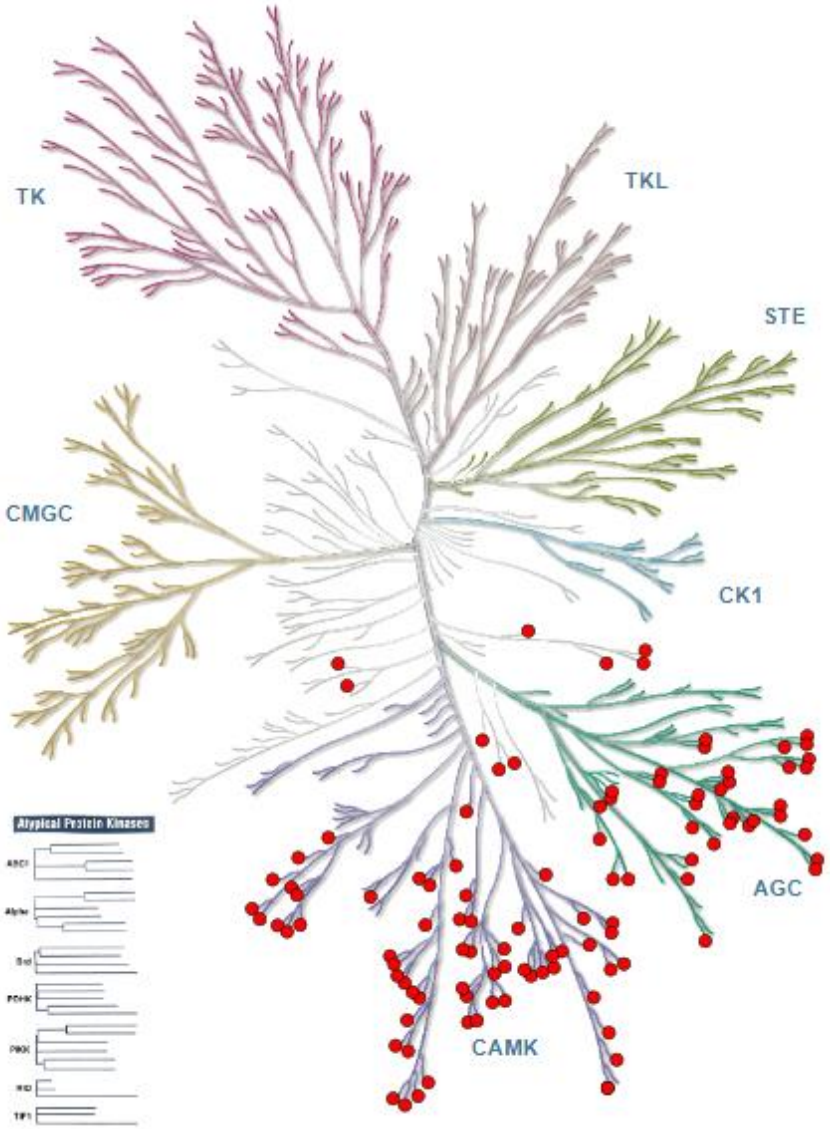
AURKB	Aurora Kinase B

PKN2	Protein Kinase N2
PRKG2	Protein Kinase G2
PRKG1	Protein Kinase G1
PKN1	Protein Kinase N1
MAST3	Microtubule-Associated Serine/Threonine Kinase 3
MAST2	Microtubule-Associated Serine/Threonine Kinase 2
PLK1	Polo-Like Kinase 1
PLK3	Polo-Like Kinase 3
STK36	Serine/Threonine Kinase 36



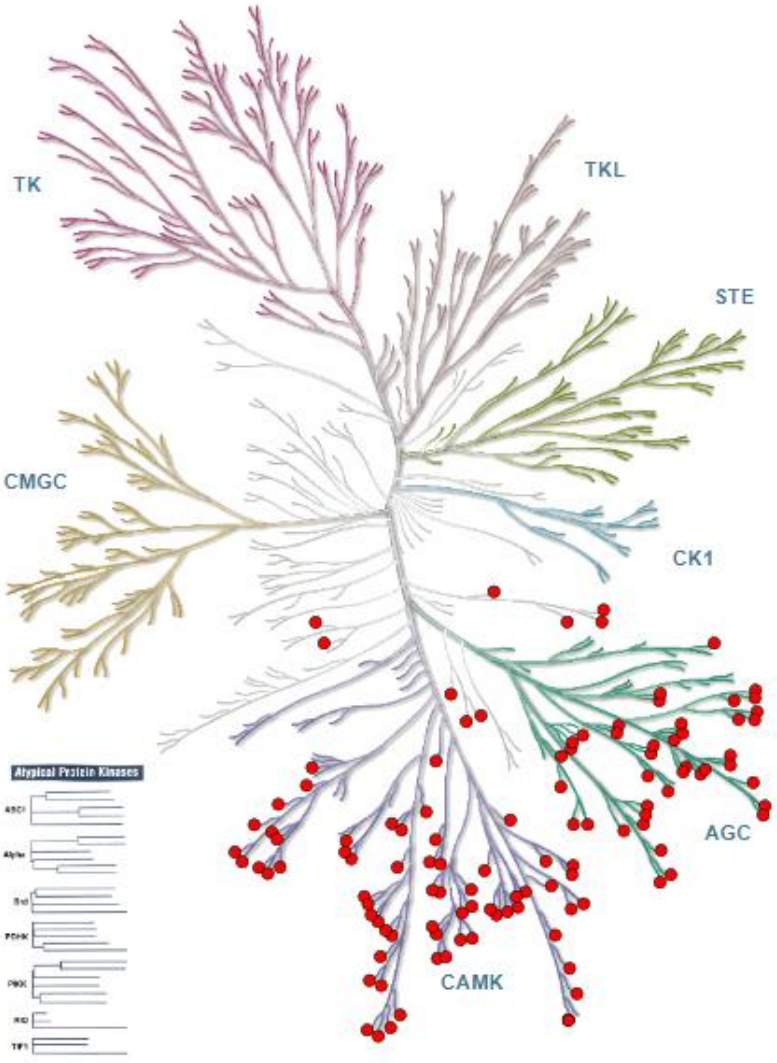
Iteration 11: Max target sequnce = 110

MAST1	Microtubule-Associated Serine/Threonine Kinase 1
MAST4	Microtubule-Associated Serine/Threonine Kinase 4
PKN3	Protein Kinase N3
PLK2	Polo-Like Kinase 2
STK38L	Serine/Threonine Kinase 38 Like
STK38	Serine/Threonine Kinase 38
DMPK	Myotonic Dystrophy Protein Kinase
TSSK3	Testis-Specific Serine Kinase 3
ROCK1	Rho-Associated Coiled-Coil Containing Protein Kinase 1
CDC42BPA	Cell Division Cycle 42 Binding Protein Kinase Alpha



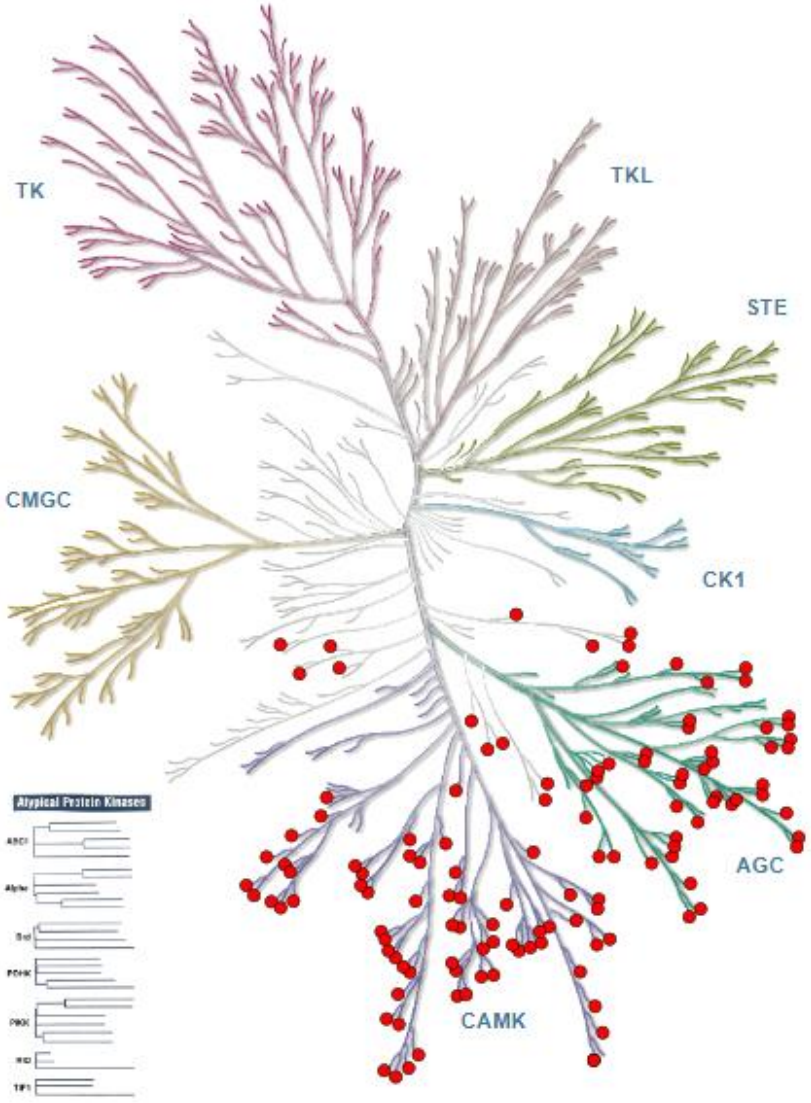
Iteration 12: Max target sequnce = 120

CDC42BPB	CDC42 binding protein kinase beta
CDC42BPG	CDC42 binding protein kinase gamma
ROCK2	Rho-associated coiled-coil containing protein kinase 2
LATS2	Large tumor suppressor kinase 2
TSSK2	Testis-specific serine kinase 2
TSSK1B	Testis-specific serine kinase 1B
CIT	Citron Rho-interacting kinase
OBSCN	Obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF
GRK5	G protein-coupled receptor kinase 5
CAMPKK2	Calcium/calmodulin-dependent protein kinase kinase 2



Iteration 13: Max target sequnce = 130

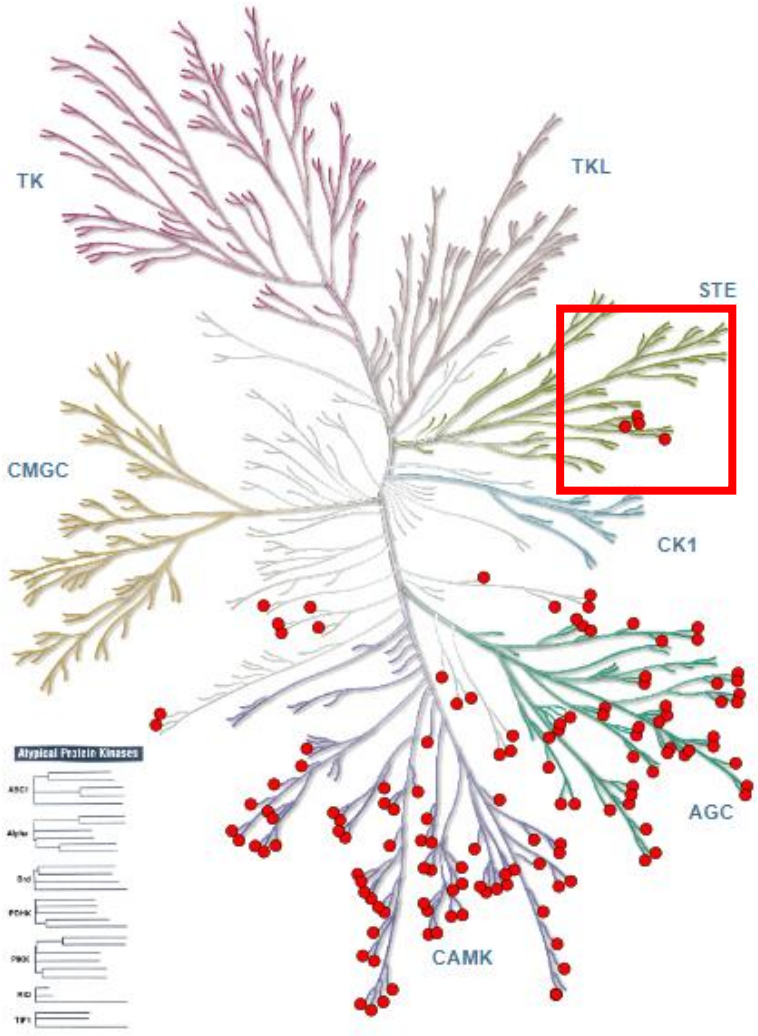
GRK6	G Protein-Coupled Receptor Kinase 6
GRK4	G Protein-Coupled Receptor Kinase 4
GRK1	G Protein-Coupled Receptor Kinase 1
TSSK4	Testis-Specific Serine Kinase 4
TSSK6	Testis-Specific Serine Kinase 6
CAMKK1	Calcium/Calmodulin-Dependent Protein Kinase Kinase 1
STK32B	Serine/Threonine Kinase 32B
ULK2	Unc-51 Like Autophagy Activating Kinase 2
STK32A	Serine/Threonine Kinase 32A
GRK3	G Protein-Coupled Receptor Kinase 3



Iteration 14: Max target sequnce = 140

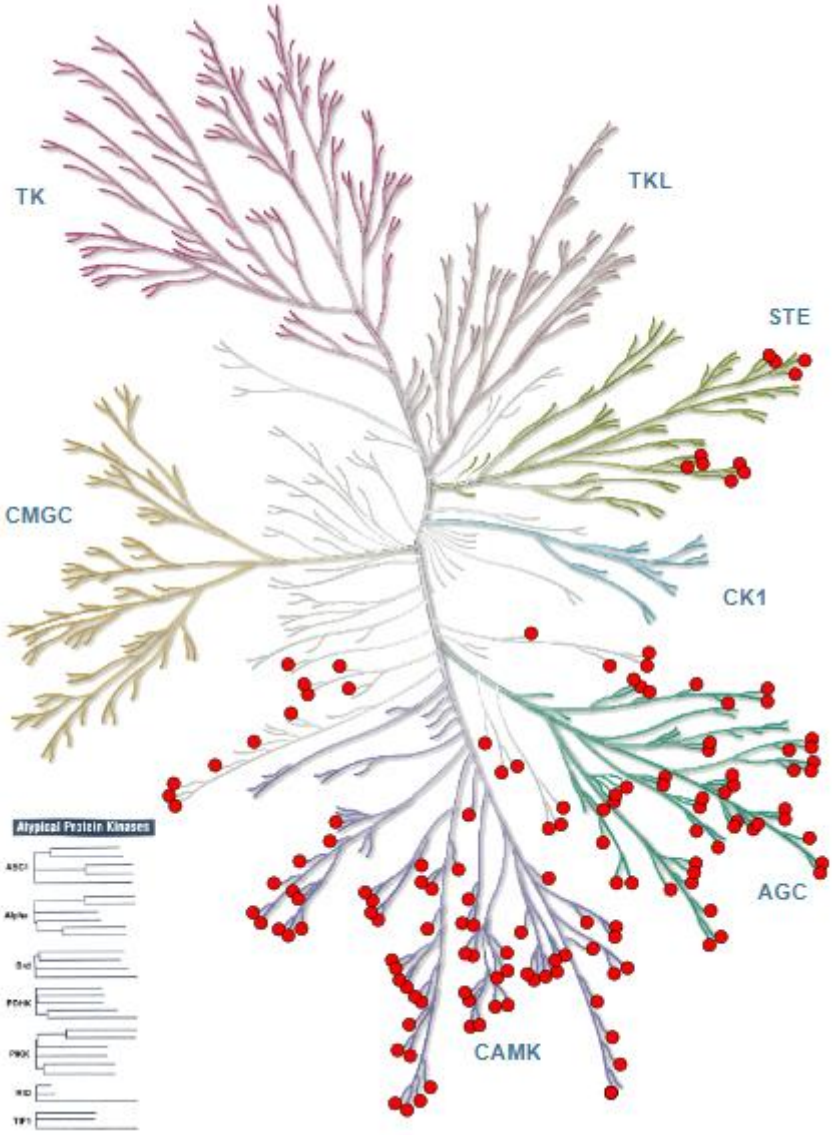
In the 14th iteration, a kinase from the STE kinase group was observed for the first time. The STE kinase group represents a specific protein kinase family called Signal Transducing and Extracellular Signal-Regulated Kinase (STE).

STK32C	Serine/Threonine Kinase 32C
GRK2	G Protein-Coupled Receptor Kinase 2
GRK7	G Protein-Coupled Receptor Kinase 7
ULK1	Unc-51 Like Autophagy Activating Kinase 1
NEK5	Never In Mitosis A-Related Kinase 5
NEK3	Never In Mitosis A-Related Kinase 3
PAK6	p21-Activated Kinase 6
PAK1	p21-Activated Kinase 1
PAK2	p21-Activated Kinase 2
NEK	Never In Mitosis A-Related Kinase
PAK3	p21-Activated Kinase 3



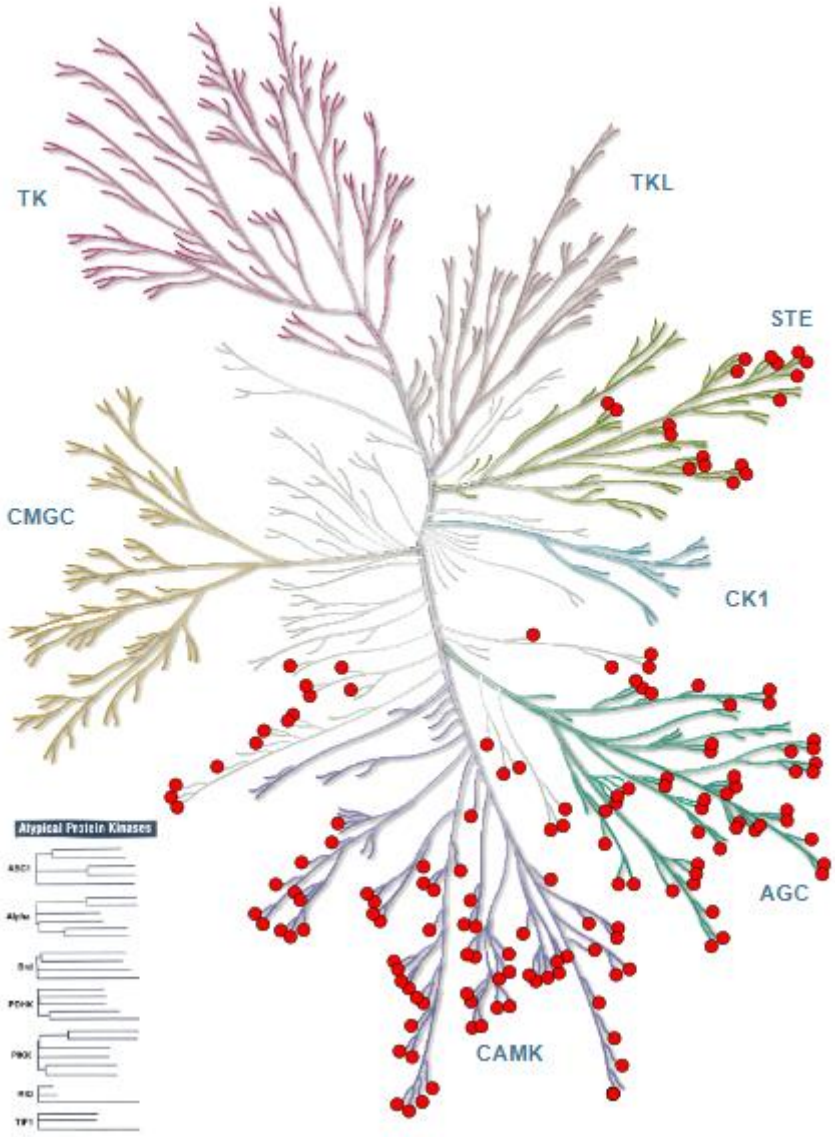
Iteration 15: Max target sequnce = 150

PAK4	p21-activated kinase 4
PAK5	p21-activated kinase 5
NEK1	NIMA (never in mitosis gene a)-related kinase 1
STK3	Serine/threonine kinase 3
STK4	Serine/threonine kinase 4
NEK4	NIMA (never in mitosis gene a)-related kinase 4
STK24	Serine/threonine kinase 24
NEK2	NIMA (never in mitosis gene a)-related kinase 2
NEK6	NIMA (never in mitosis gene a)-related kinase 6
STK26	Serine/threonine kinase 26
PAK4	p21-activated kinase 4



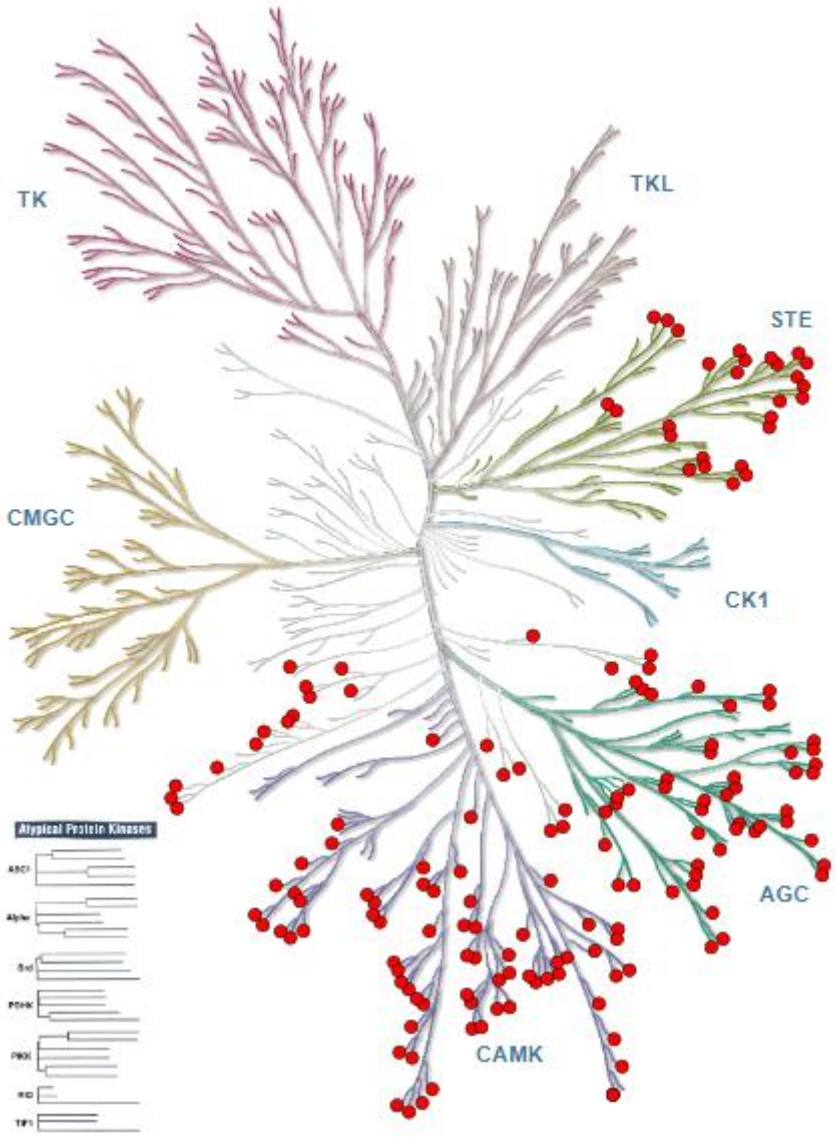
Iteration 16: Max target sequnce = 160

STK25	Serine/Threonine Kinase 25
NEK7	NIMA (Never In Mitosis Gene A)-Related Kinase 7
MAP4K3	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 3
OXSR1	Oxidative-Stress Responsive 1 Kinase
SLK	STE20-like Kinase
NEK9	NIMA (Never In Mitosis Gene A)-Related Kinase 9
MAP4K2	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 2
STK10	Serine/Threonine Kinase 10
STK39	Serine/Threonine Kinase 39
TNIK	TRAF2- and NCK-interacting Kinase



Iteration 17: Max target sequnce = 170

MAP4K5	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 5
MAP4K1	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 1
MINK1	Misshapen/NIK-related kinase 1
MYO3B	Myosin III B
MAP3K3	Mitogen-Activated Protein Kinase Kinase Kinase 3
MAP4K4	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4
MYO3A	Myosin III A
CHEK1	Checkpoint Kinase 1
MAP3K5	Mitogen-Activated Protein Kinase Kinase Kinase 5
MAP3K2	Mitogen-Activated Protein Kinase Kinase Kinase 2

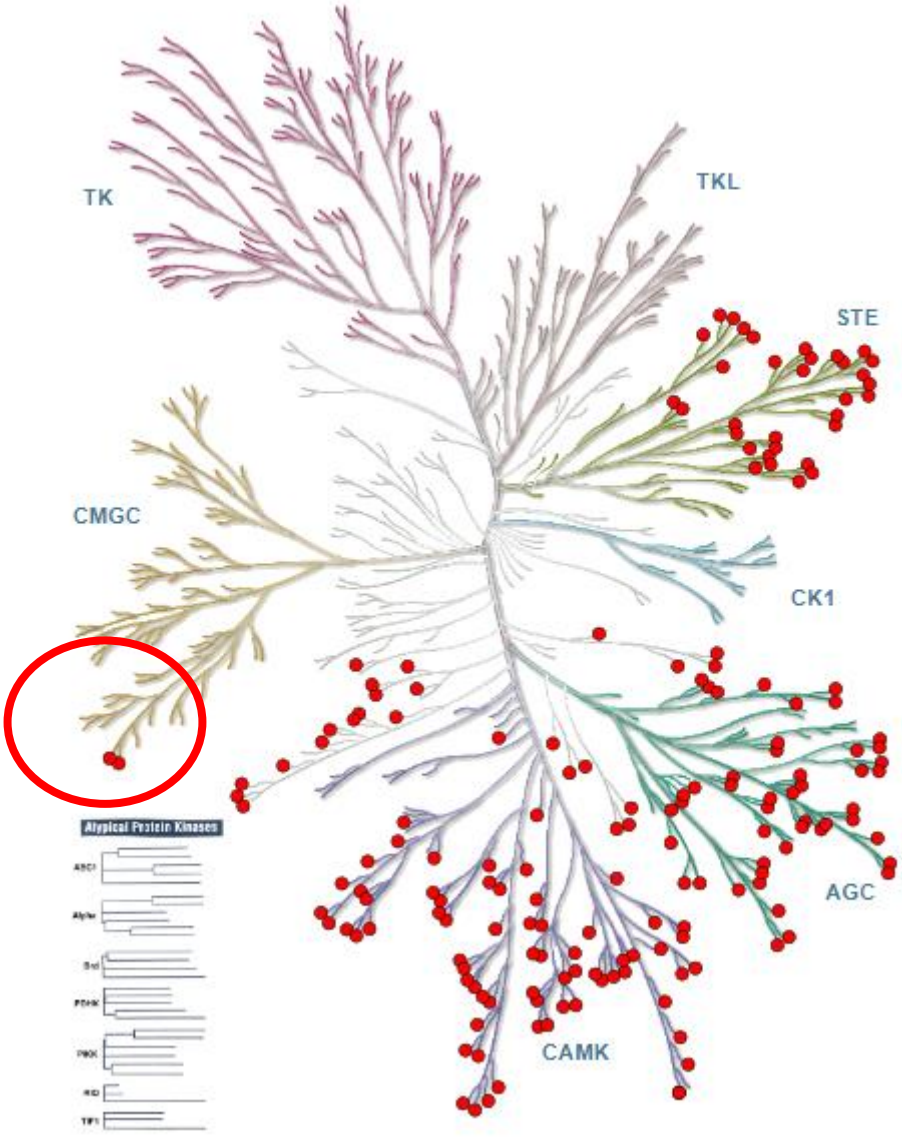


Iteration 18: Max target sequnce = 180

In the 18th iteration, a kinase from the 4th distinct kinase group was observed. According to the given instructions, we are required to stop the iteration when three groups are filled, and a kinase is marked for the first time in the 4th group. In this iteration, after the CAMK, AGC, and STE kinase groups were filled, the iteration ended as a kinase was marked for the first time in the CMGC group.

The CMGC kinase group represents a protein kinase family that encompasses a wide biological spectrum, including various kinases such as Cyclin-Dependent Kinase (CDK), Mitogen-Activated Protein Kinase (MAPK), Glycogen Synthase Kinase (GSK), and Cyclin-Dependent Kinase-Like (CDKL), involved in processes ranging from cell cycle regulation to cellular signal transduction.

MAP3K6	Mitogen-Activated Protein Kinase Kinase Kinase 6
MAP3K15	Mitogen-Activated Protein Kinase Kinase Kinase 15
MAP3K19	Mitogen-Activated Protein Kinase Kinase Kinase 19
TAOK1	Thousand and one amino acid kinase 1
TAOK2	Thousand and one amino acid kinase 2
CDK2	Cyclin-Dependent Kinase 2
TAOK3	Thousand and one amino acid kinase 3
KALRN	Kalirin, RhoGEF Kinase
CDK3	Cyclin-Dependent Kinase 3
NEK10	NIMA (Never In Mitosis Gene A)-Related Kinase 10


TK
TKL
STE
CMGC
CK1
AGC
CAMK
Atypical Protein Kinases