Phylogenetic Evaluation of Potential ERK Inhibitors for Lung Cancer Treatment

Rayden Yap, Ariana Budhihartanto, Fiona Hiu, Olivia Kandila, Geraldus Flabian

1) 1st Author Affiliation, Jakarta Intercultural School, Jakarta, Indonesia (1rayden.yap@gmail.com)
2) 2nd Author Affiliation, Jakarta Intercultural School, Jakarta, Indonesia (arianajasminb@gmail.com)
3) 3rd Author Affiliation, NationalHigh Jakarta School, Jakarta, Indonesia (hiu.fiona@gmail.com)
4) 4th Author Affiliation, Bina Bangsa School PIK, Jakarta, Indonesia (olivia.kandila@gmail.com)
5) 5th Author Affiliation, SMAK 1 Penabur, Jakarta, Indonesia (geraldus.flabian@gmail.com)

Abstract

All cells experience changes through biological pathways, with the mitogen-activated protein kinase (MAPk) pathway, also known as the Ras-Raf-MEK-ERK pathway, being responsible for regulating proliferation and apoptosis. The ERK target gene regulates cellular functions that are most commonly used by cancer cells to excessively proliferate. The organisms Vitis vinifera, Camellia sinensis, Angelica gigas, Ferula assafoetida, Cinnamomum cassia, Duchesnea indica, and Psidium cattleianum were discovered to contain secondary metabolites with inhibitory properties. Phylogenetic analyses using distance-based and character-based methods are used to identify evolutionary relationships between species that produce said secondary metabolites. Species were individually analyzed through internal transcribed spacer 2 (ITS2) nucleotide sequences where Vitis vinifera and Angelica gigas were the most closely related to each other. This study recommends that the evolutionary relationship between Vitis vinifera and Angelica gigas be utilized for the mass production of lung cancer treatment, in addition to the exploration of molecular docking to understand how each secondary metabolite inhibits the ERK pathway.

Keywords: *phylogenetic analysis, ERK 1/2, cancer cell, lung cancer, ITS2 sequence*

1. Introduction

All cells experience changes through biological pathways that regulate cellular processes such as apoptosis, differentiation, and proliferation. Cell proliferation involves the growth of the cell and the division of its DNA by splitting into two daughter cells while apoptosis describes a programmed cell death that self-destructs cells through chromatin condensation, nuclear fragmentation, and plasma membrane blebbing (1, 2).

Each biological pathway occurs through the initial transmission of intermolecular signals from the receptors on the surface of a cell to the DNA inside its nucleus. Genes cascade within the cell where select concentrations of substrates are phosphorylated, or expressed, into the nucleus of a cell. Consequently, transcription occurs and cells are able to carry out biological functions.

The mitogen-activated protein kinase (MAPk) pathway, also known as the Ras-Raf-MEK-ERK pathway, is responsible for regulating proliferation and apoptosis. It is initiated by ligands binding to receptor tyrosine kinases found inside the cell, which activates a kinase cascade that allows a chain of proteins to dock onto one another. Figure 1 shows that epidermal growth factor (EGF) ligands outside the cell bind to EGF receptors in the cell membrane

to cause dimerization (3). A protein kinase cascade involving the rat sarcoma virus (RAS), rapidly accelerated fibrosarcoma (RAF), mitogen-activated protein (MEK), and extracellular signal-regulated (ERK) kinases are activated (4, 5).

Cancer cells are a direct result of damaged or mutated genes that express different concentrations of substrates into the DNA of a cell, leading it to carry out dysregulated transcriptions (6). In certain cases, transcriptions due to genetic mutation allow the creation of more cancer cells. As cancer cells accumulate, they invade surrounding tissues for survival, form tumors, and spread to new locations in the body through a process known as metastasis.

The ERK kinase phosphorylates gene expressions that regulate cell cycle checkpoints in response to extra- and intracellular signals, such as the cyclin-dependent kinase inhibitor (CDKI) gene (7), and cellular transcription through the interaction with activator proteins, such as the fos and jun proto-oncogenes (8).

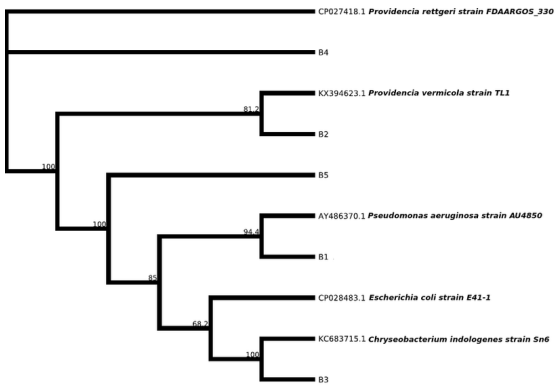
Mitogen-activated protein kinase (MAPK) pathways play a key role in signaling cascades that regulate a variety of cellular processes including proliferation, differentiation, apoptosis, and stress responses.

There are four MAPK cascades identified in eukaryotic cells: the extracellular signal-regulated kinase (ERK) cascade, the p38MAPK cascade, the c-Jun N-terminal kinase cascade, and the ERK5 cascade. While the p38MAPK and c-Jun N-terminal cascades are mainly involved in transducing stress-related stimuli, the ERK cascade is mainly involved in the transmission of mitogenic signals. ERK is a type of serine/threonine protein kinase, usually known as ERK1/2. The ERK cascade or RAS-RAF-MEK-ERK pathway is the most well-studied of the MAPK cascades and is important for cell growth, differentiation, and survival. Activation of receptor tyrosine kinases by growth factors and various extracellular signals leads to the sequential activation of RAS, RAF, MEK, and ERK, which is generally located in the cytoplasm, will enter the nucleus and regulates transcription factor activity and gene expression (9).

The ERK/MAPK signaling pathway is not only involved in regulating cellular biological functions, such as cell proliferation, cell differentiation, cell cycle regulation, cell apoptosis and tissue formation, but is also related to tumor formation. Tumor formation includes aberration of biological processes such as unlimited cell proliferation, dedifferentiation and a lack of apoptosis. The activation of the ERK/MAPK signaling pathway promotes proliferation and has an anti-apoptotic effect.

Continuous activation of the ERK/MAPK signaling pathway can promote the transformation of normal cells into tumor cells, while inhibition of the ERK/MAPK signaling pathway can restore tumor cells to a non-transformed state *in vitro* and can inhibit tumor growth *in vivo*. Therefore, increased activation of the ERK/MAPK signaling pathway may be closely related to the occurrence and development of tumors.

Scientific disciplines such as phylogeny provide new insight into controlling cancer cell evolution and effectively treating disease. Phylogenetic analysis analyzes the changes occurring in different organisms during evolution and the biological ancestors a group of organisms may share (10). Methods like genetic comparison and multiple sequence alignment are used to form phylogenetic trees as shown in figure 2: each branch represents the relationship between organisms with respect to ancestry and descent while the branch length represents the evolutionary changes that occurred in that branch.



Phylogenetic analyses are performed with the compilation of conserved nucleotide sequences, which refer to a DNA sequence that remains unchanged throughout an organism's evolution. Identical nucleic acid, polymeric carbohydrate, or protein sequences can be produced across species, indicating that multiple species conserved sequences after evolution. Given that genetic information is transmitted to offspring, a shared conserved sequence

indicates whether species were of the same descent and contained identical genes (11).

A prominent conserved sequence used for phylogenetics is the internal transcribed spacer 2 (ITS2) nucleotide region, which allows for the identification of closely related species within families and genera (12). This study will be utilizing ITS2 sequences to draw conclusions on how inhibitors of the ERK pathway are related to each other to optimize their use for lung cancer treatment.

Applying phylogenetic principles to lung cancer research promotes the understanding of how cancer cells mutate, along with identifying species with inhibitory properties for said mutations according to how genetically related they are. Phylogenetic trees are created with distance-based methods, such as neighbor-joining and the Unweighted Pair Group Method with Arithmetic (UPGMA), or character-based methods, such as Bayesian interference, the maximum likelihood method, and maximum parsimony methods.

New immune checkpoint inhibitors were recently approved for lung cancer treatment but are ineffective because of the rate at which cancer cells become resistant to chemotherapeutic agents. Previous literature indicates the potential of certain bioactive substances as checkpoint inhibitors for the ERK pathway that are also naturally obtained. By using phylogenetic analysis, the characteristics of said bioactive substances such as common ancestors and geographical location can be applied to source inhibitory treatment for lung cancer. This study aims to identify the evolutionary relationships between *Vitis vinifera*, *Camellia sinensis*, *Angelica gigas*, *Ferula assafoetida*, *Cinnamomum cassia*, *Duchesnea indica*, and *Psidium cattleianum* as secondary metabolite sources with the potential to inhibit the ERK pathway.

2. Method and Experimental Details

Data Acquisition

In silico phylogenetic analysis was conducted to identify the evolutionary relationships between secondary metabolites through Internal Transcribed Spacer 2 (ITS2) sequence variability. Representative species of each metabolite were obtained from online data sets; wherever possible, at least ten species of each taxon were used. Table 1 shows a summary of contemporary research on organisms with secondary metabolites that inhibit the expression and mechanisms of the ERK target gene.

Secondary Metabolites	Inhibition Mechanism	Organisms
Pterostilbene	Decreases epidermal growth factor expression and induces apoptosis by decreasing levels of Caspase-3 and LC3-II Chen et al., 2012	*Vitis vinifera, Vaccinium cyanococcus, Vaccinium oxycoccus*
Epigallocatechin-3-gallate	Suppresses the invasion of carcinomic A549 cell lines Xue et al., 2015	*Camellia sinensis*
Decursin	Inhibits the phosphorylation of the p42, p44, ERK, and JNK target genes Jung et al., 2009	*Angelica gigas*
Farnesiferol C	Inhibits the phosphorylation of the p125 FAK (pY861), Human Src (pY416) Lee et al., 2010, ERK 1/2, p38MAPK, and JNK target genes	*Ferula assafoetida*
C. cassia Extract	Inhibits the phosphorylation of the FAK and ERK1/2 target genes and reduces the metastasis of adenocarcinoma cells Wu et al., 2018	*Cinnamomum cassia*
D. indica Extract	Inhibits the phosphorylation of the p-ERK, p-FAK Tyr397, p-paxillin Tyr118, c-Jun, c-Fos, and TGF-b1 induced-vimentin target genes Chen et al., 2017	*Duchesnea indica*
Butanol Fraction of *P. cattleianum* Extract	Suppresses the expression of the matrix metalloproteinases (MMP)-9 and MMP-2 target genes by downregulating the ERK 1/2 pathways Im et al., 2012	*Psidium cattleianum*

Data Preprocessing

Computer-aided multiple sequence alignment of the ITS2 sequences was performed with Clustal Omega, a web-based software from the EMBL European Bioinformatics Institute. 47 ITS2 sequences were aligned using distance-based methods, which included the *Vitis vinifera*, *Camellia sinensis*, *Angelica gigas*, *Aspergillus wentii*, and *Phyllanthus emblica* species. A portion of the secondary metabolites highlighted in table 1 were not used for the analysis due to the lack of sequencing data available. The analysis was conducted with the following parameters: 1) input sequence alignment, 2) generating Mbed-like clustering guide trees, 3) Mbed-like clustering iterations, 4) default combined iterations, 5) default maximum guide tree iterations, 6) default maximum Hidden Markov Model (HMM) iterations, 7) no distance matrices, 8) guide tree generation, and 9) aligned orders. All input sequences were stored as a NEXUS file.

Preprocessed data was also analyzed using PhyloML, a web-based software from the PACA Bioinfo platform (13) that utilizes the SH-like Approximate Likelihood-Ratio Test (aLRT) method to form phylogenetic trees. The General Time-Reversible (GTR) substitution model was utilized with the following parameters: 1) 4 substitution rate categories, 2) estimated gamma distribution parameters, 3) estimated proportion of invariable sites, and 4) estimated transversion ratios. A phylogenetic tree was also generated with the use of the web-based Interactive Tree Of Life (iTOL) platform (14).

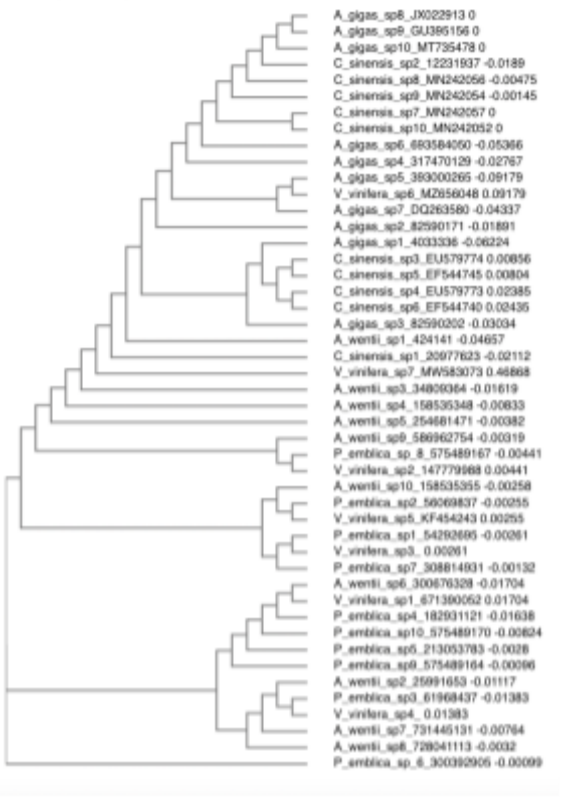


Figure 3: a cladogram generated through distance-based methods does not depict evolutionary distance between organisms as the branch length does not change.

3. Results and Discussion

Both distance-based and character-based phylogenetic trees were generated as illustrated in figures 3-5.

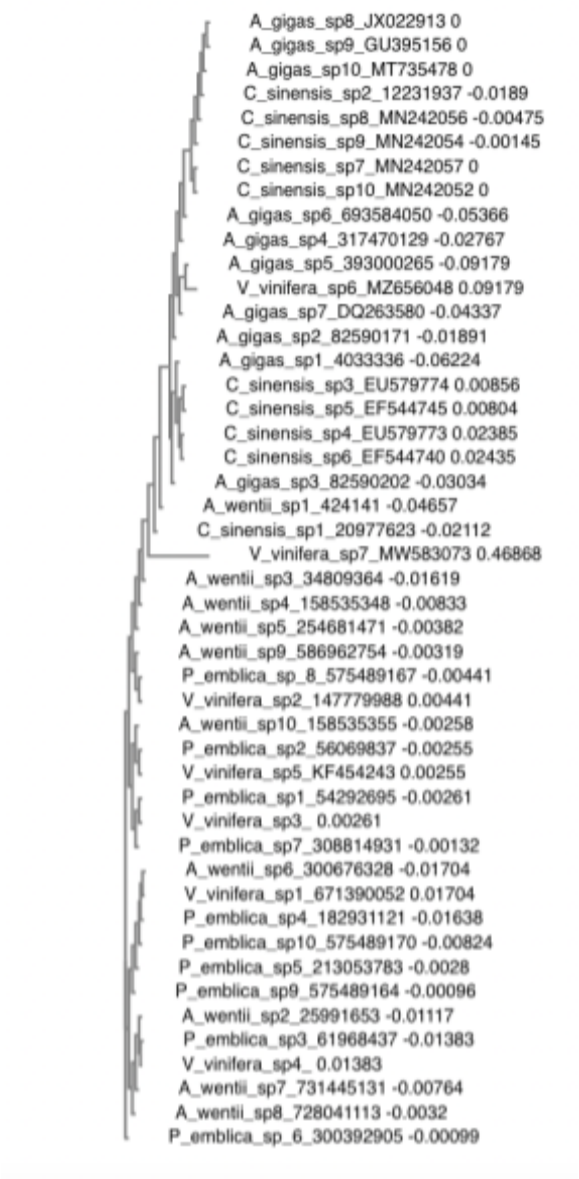


Figure 4: a phylogenetic tree generated through distance-based methods contains varying branch lengths when comparing two or more sequences.

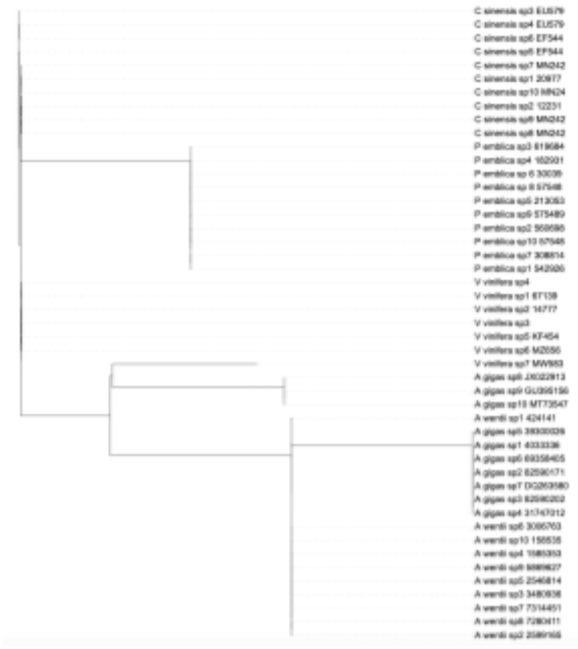


Figure 5: a phylogenetic tree generated through character-based methods allows for the identification of outliers, which are distantly related taxa that are not similar to other sequences.

Distance-based methods calculate the overarching differences between organisms and construct a distance matrix to depict how closely or distantly related organisms are to each other. Based on the distance scores calculated, a phylogenetic tree is generated. Construction methods such as the neighbor-joining and UPGMA methods are advantageous in the sense that they are more time-efficient and widely accessible as opposed to character-based methods. In addition, the UPGMA is reliable in identifying related sequences, which capture important similarities among species. However, their efficiency leads to a loss of information when converting between sequence data to distance data, potentially producing incorrect results.

Character-based methods detect individual sequences, or characters, by assuming that each character is independent of its neighbors: therefore, phylogenetic trees are predicted to show the smallest amount of evolutionary changes to distinguish two organisms from one another. Character-based phylogenetic trees are comparatively more accurate and are more useful in forming hypotheses on evolutionary relationships. Despite this, construction methods such as maximum likelihood and maximum parsimony are only appropriate for smaller datasets and are susceptible to slow search algorithms.

are only appropriate for smaller datasets and are susceptible to slow search algorithms.

As shown in figure 5, the branch lengths of organisms vary according to their defined existence. *Camellia sinensis* species 3-6 are outliers and are not directly related to the other 4 species. However, both the *Vitis vinifera* and *Angelica gigas* species are monophyletic, or closely related to each other - with a thorough understanding of the climates in which both species grow in, artificial environments may be created for them to grow together. The *Vitis vinifera* and *Angelica gigas* species are both classified under the asterid clade, hence sharing the same environmental conditions for growth. Each species requires deep, moist, and loamy soil in full or partial shade (15, 16).

4. Conclusion

As species that produce secondary metabolites with inhibitory activity, Vitis vinifera and Angelica gigas can be culturalized for mass production to treat future lung cancer cases. However, noting the limitations of this study, the investigation of the species' inhibitory properties can be explored through molecular docking.

5. Acknowledgments

This study thanks Mr. Juan Bahrian and Mr. Paul Edison for the immense support provided, Sebastien Santini (CNRS/AMU IGS UMR7256), and the PACA Bioinfo platform for the availability and management of the phylogeny.fr website used to process data on ITS2 sequences to perform character-based analysis.

6. References

[1] Shackleton, M. (2010). Normal stem cells and cancer stem cells: similar and different. Seminars in Cancer Biology, 2, 85–92. https://doi.org/10.1016/j.semcancer.2010.04.002
[2] Kerr, J. F. R., Wyllie, A. H., & Currie, A. R. (1972). Apoptosis: A Basic Biological Phenomenon with Wideranging Implications in Tissue Kinetics. British Journal of Cancer, 4, 239–257. https://doi.org/10.1038/bjc.1972.33
[3] Wu, P.-K., Becker, A., & Park, J.-I. (2020). Growth Inhibitory Signaling of the Raf/MEK/ERK Pathway. International Journal of Molecular Sciences, 15, 5436. https://doi.org/10.3390/ijms21155436
[4] Eblen, S. T. (2018). Extracellular-Regulated Kinases: Signaling From Ras to ERK Substrates to Control Biological Outcomes. In Advances in Cancer Research (pp. 99–142). Elsevier. http://dx.doi.org/10.1016/bs.acr.2018.02.004
[5] McKay, M. M., & Morrison, D. K. (2007). Integrating signals from RTKs to ERK/MAPK. Oncogene, 22, 3113–3121. https://doi.org/10.1038/sj.onc.1210394
[6] The cancer genome | Nature. (n.d.). Nature; Springer Nature. Retrieved October 18, 2022, from https://doi.org/10.1038/nature07943
[7] Ding, L., Cao, J., Lin, W., Chen, H., Xiong, X., Ao, H., Yu, M., Lin, J., & Cui, Q. (2020). The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. International Journal of Molecular Sciences, 6, 1960. https://doi.org/10.3390/ijms21061960
[8] CURRAN, T. (1992). Fos and Jun. Oncogenic Transcription Factors. The Tohoku Journal of Experimental Medicine, 2, 169–174. https://doi.org/10.1620/tjem.168.169
[9] Barbosa, R., Acevedo, L. A., & Marmorstein, R. (2021). The MEK/ERK Network as a Therapeutic Target in Human Cancer. Molecular Cancer Research, 3, 361–374. https://doi.org/10.1158/1541-7786.mcr-20-0687
[10] Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. Nature Reviews Genetics, 7, 428–444. https://doi.org/10.1038/s41576-020-0233-0
[11] Dermitzakis, E. T., Reymond, A., & Antonarakis, S. E. (2005). Conserved non-genic sequences — an unexpected feature of mammalian genomes. Nature Reviews Genetics, 2, 151–157. https://doi.org/10.1038/nrg1527
[12] Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., & Chen, S. (2010). Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. PLoS ONE, 10, e13102. https://doi.org/10.1371/journal.pone.0013102
[13] Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W465-9. Epub 2008 Apr 19.
[14] Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research, W1, W293–W296. https://doi.org/10.1093/nar/gkab301
[15] Topalović, A., Knežević, M., Bajagić, B., Ivanović, L., Milašević, I., Đurović, D., Mugoša, B., Podolski-Renić, A., & Pešić, M. (2020). Grape (Vitis vinifera L.): health benefits and effects of growing conditions on quality parameters. In Biodiversity and

Biomedicine (pp. 385–401). Elsevier. http://dx.doi.org/10.1016/B978-0-12-819541-3.00020-7

[16] Park, Y., Park, P. S., Jeong, D. H., Sim, S., Kim, N., Park, H., Jeon, K. S., Um, Y., & Kim, M.-J. (2020). The Characteristics of the Growth and the Active Compounds of Angelica gigas Nakai in Cultivation Sites. Plants, 7, 823. https://doi.org/10.3390/plants9070823