Asian Journal of Convergence in Technology **Volume V, Issue I**
ISSN No.:2350-1146, I.F-5.11

Prediction and Analysis of Allergenic Epitopes of Tree-Nuts and its Cross-Reactivity

Amogh Johri
International Institute of Information Technology, Bangalore
Bengaluru, India
amoghjohri@yahoo.com

Meenakshi Srivastava
Amity Institute of Information Technology
Lucknow, India
msrivastava@lko.amity.edu

***Abstract*—With a whopping 10-25% of the world's population being affected by it, allergies have become one of the top 10 reasons for visit to primary care physicians. Among this, tree-nut allergies are one of the most common allergies causing food substances. In the contemporary times various computational tools have emerged in order to facilitate time and cost-effective study of food allergens. This does not only aid in fabrication of a cure but also in its prevention as by analyzing for cross-reactivity among different allergens, patients can be advised against a number of possible other food substances which are likely to trigger the same response by their immune system. In the present study, the tool being utilized is EpiPro1.0 which has been developed by the authors in order to carry out accurate and efficient epitope prediction of an allergenic sequence (FASTA format). The tool also utilizes a novel algorithm in order to find the consensus of the results obtained through a number of different web-servers. In the present study, 20 different allergenic sequences from 6 major allergy causing tree-nuts, namely Almonds, Black Walnut, Brazil Nut, Cashew Nut, English Walnut and Hazel Nut, have been analyzed and 326 possible allergy causing epitopes have been predicted. Since, patients suffering from one tree-nut allergy tend to show sensitivity towards other tree-nuts as well, their cross-reactivity has also been studied in order to make accurate predictions regarding possible allergic reactions.**

Keywords—Allergens, B-cell epitopes, consensus, cross-reactivity

I. INTRODUCTION

Even in the contemporary age of technological development, allergies still continue to pose a daunting challenge for the medical society [1,2]. This has provided for a number of open-source computational tools to be developed in order to carry out time and cost-efficient predictions and analysis. The primary step in allergy analysis is to obtain B-cell epitopes of the allergenic sequence, which appear as amino-acid subsets. Chemically these allergens are proteins (chain of amino-acids) and it is due to certain subsets of these amino-acids which renders a protein its allergenic properties [3].

Obtaining these epitopic sequences from its parent protein is a daunting task in itself. Experimental methods for carrying out epitope prediction are often very costly and time inefficient. In addition, they also struggle in terms of accuracy of their results[4]. Hence, number of computational tools are being developed in order to perform epitope prediction. Researchers have emphasized that if a consensus can be obtained of the results predicted through different web-servers it would have a significant positive impact over the accuracy of the so obtained result [5]. This served as the motivation for developing EpiPro1.0[6] (Epi(tope) Pro(gnosis)1.0), a novel tool developed by the author in order to make up for the various drawbacks and carry out simple, efficient and a more accurate epitope prediction and analysis. EpiPro1.0 is an opensource software available at (https://github.com/amogh7/EpiPro1.0) which has been developed on python2.7 and html5. It provides as an umbrella in order to incorporate a number of epitope prediction web-servers based over different algorithms (random forest algorithm, recurrent neural networks, etc) and facilitates a bio-informatician to carry out 'one-click' epitope prediction through each of the individual web-servers without having to go through the understanding of their different interfaces. In addition, EpiPro1.0 also provides the user with a novel algorithm in order to get the consensus from the results obtained through different web-servers. While obtaining the consensus is the dominant advantage provided by the tool, leaving it out still renders a massive increase in efficiency, by almost ~1650% in comparison to the current methodologies in place.

In the present study, analysis has been carried out over tree-nut allergens. These are one of the most prominent food allergies, in terms of both the number of people affected and the severity of the allergic reaction it induces. In addition to being highly unpredictable, even very small amounts of tree nuts are deemed potent enough to induce an allergic reaction. Furthermore, the high probability of its patient going into anaphylactic shock demands a patient, allergic to tree-nuts, to carry epinephrine auto-injector at all times[7]. Hence, an extensive analysis of tree-nut allergens shall be of immense aid to the medical society.

In the present study, authors have moved one step further to develop EpiPro1.1 which also enables the user in order to analyze the results of different allergens to be compared for cross-reactivity. As patients allergic to one type of tree-nut often show high levels of sensitivity to a number of other tree-nuts, extensive analysis of cross-reactivity between different tree-nuts is of major importance. In the present study, EpiPro1.1 has been utilized in order to carry this out and a major number of hits have been detected.

The present paper incorporates the study of six different tree-nuts, namely Almonds(Prunus dulcis), Black Walnut(Jutland Nigar), Brazil Nut(Bertholletia excelsa), Cashew Nut(Anacardium occidentale), English Walnut(Jutland regia) and Hazelnut(Corylus). A total of twenty allergenic sequences belonging to one of the six mentioned tree nuts were analysed, namely : Almonds - 'Pru du 3', 'Pru du 4', 'Pru du 5' and 'Pru du 6'; Black Walnut - 'Jug n 1' and 'Jug n 2'; Brazil Nut - 'Ber e 1' and 'Ber e 2'; Cashew Nut - 'Ana o 1', 'Ana o 2' and ' Ana o 3'; English Walnut - 'Jug r 1', 'Jug r 2', 'Jug r 3' and 'Jug r 4';Hazelnut - 'Cor a 1'. 'Cor a 8', 'Cor a 9', 'Cor a 11', and 'Cor a 14'.

II. STATE OF ART

In the modern age, researchers are equipped with a number of computational tools to facilitate epitope prediction. These

tools, being based over different algorithms often show discrepancy amongst their results. A number of studies have suggested that a consensus of the results obtained from multiple such tools shall significantly bolster the accuracy. However, to the best of the knowledge of the authors, no such tool could be found in literature which would render the facility of obtaining the consensus. In their previous study authors have developed a tool to predict the epitopes using various web-servers. A novel algorithm has also been developed, which would provide EpiPro1.0 to obtain the consensus from the results of a number of different web- servers (six in the present case).

III. Materials and Method

The allergenic protein sequences(FASTA format) were obtained from National Centre of Biotechnology Information(www.ncbi.nlm.nih.gov)[8].

A total of six web-servers were used through EpiPro1.0 in the present study, namely: AAP[9], FBCPred[10], BcePred[11], BCPred[12], BepiPred2.0[13] and ABCPred[14].

IV. Result and Discussion

The various allergenic sequences of the six tree-nuts were analyzed and their possible epitopes were predicted using the consensus result of the six different web-servers (as mentioned efore). The length of the predicted epitopic sequences is also of importance (the longer the sequence, the more likely it is to be epitopic) and EpiPro1.0 allows the user to obtain consensus predictions of desired lengths [6]. Here, the length of the sequences considered in the result is equal to or greater than 5. Table I. summarizes the result.

TABLE I. Predicted epitopic sequences

Tree-Nut	Allergen	Predicted Sequences
Almond	Pru du 3	PKAMAA
		TPCINYVAN
		NYVANGGALN
		IPYKISPST
		VNGIPYTNANAG
		ALNPSCCTG
	Pru du 4	GAVIRGKKGS
		YDEPLTPGQCN
		DQPGTL
		PLTPGQCN
	Pru du 5	AEPKKEEKVEEKED
		PGAGAGAAAPAAAEPK
	Pru du 6	ETFEDSQ
		QQFRPSRQEGGQG
		FQGEDQQ
		FNPQQQGRQQ
		NLQGQDDNR
		QREREEKQREQEQ
		EEKQREQEQQGGGGQD
		SAGGRGDQERQQEEQQ
		PFSRSAGGRGDQERQQ
		QEEQQSQRE
		GGRGDQERQQEEQQSQ
		DFVSPF
		DFYNPQG
		QVVNENG
		ISFRTDE
		YNRQES
		LSATSPPR
		GGGGQDNGVE
		RTDENGFTN
		AETFEDSQPQ
		PPRGR
		RPSRQEGGQGQ
		NLQGQDDNRN
		SWNPSDPQFQ
		QQQQQQGQQGNGNN
		FYNPQGG
		QLNQLEARE
		ENGDPI
		PQDEFNPQQQGR
		QQQGRQQQ
		GVTESWNPS
		QQQQQQQQF
		DDNRNEIVR
		QNKEWQLNQL
		FGQNKEWQL
Cash Cashew-Nut	Ana o 1	RQYDEQQK
		EKKGREREHEEEE
		EKYYKEKKGR
		RQCERQEGGQQ
		DEDEDEDEAEEED
		HNYKREDDED
		ENKRES
		YIANNDEN
		VFHGPGGENP
		KLFEKQDQ
		MSRRGEG
		WPFTEESTGS
		LFKKDPSQS
		ANITK
		ATVAS
		KVMEKEA
		KRESIN
		HGPGGENPESFYRA
		AERIDYPPL
		GGMSVPFYNSRATK
		EHEEEEEEWGTGG
		HLSSSKSSHP
		SQCMRQCE
		SQSNKYGQ
		DEDEDEAEEEDE
		EAEEEDENP
		CMRQCERQEGGQ
		KERGQHNYKREDD
		RRGEGPKIWP
		PYVFEDEDFT
		EEEDENPYVFEDE
		SSSKSSHPSYKK
		KVQRQYDEQQKEQ
		GHFEVFHGP
		INVRQ
		EGPKIWPFTEESTGS
		RCQERYKKERGQ
		QERYKKERGQHNY
		VKECEK
		PEWRKEKEGR
		YVFEDEDFTT
		KVMEKEAKELA
		ESFYRA
		VPFYNSRATK
		RGEGPKIWPFTEES
		PQQGRQQGQS
		CYNEGN
		NQLDRTP
		DVSNSQNQ
		QQQQHQSRGR
		RGSESEEESEDEKR
		ESEDEKRRWGQRD
		SQSERGSESEEES
		PSRSQSERGSES
		LRVIRPSRSQSERGS
		ISFKTND
		SREDA
		TTLTSGES
		TMRLKENIN
		MTGISYPGCPE

	Ana o 2	CPETYQ
		VEAWDP
		EGNSPVVTVT
		LEPDNRVEYEAGTVE
		PARADIYTPEV
		NNQQTT
		RKIKFNNQ
		PARADIYTP
		DVFQQQQQHQ
		QQQQQHQSRGRNL
		AGRTS
		ALEPDNR
		QEWQQQDECQ
		DGEVREGQM
		GEGMTG
		AMTSPLAG
		QTTLTSG
		GESSH
		QRDNGIEETI
		NPKDVFQQQQQ
	Ana o 3	VEVEEDSGREQ
		VKQEVQRGGRYNQRQES
		LQQQEQIKGEEVRELYE
		IVEVEEDSGRE
Brazil-Nut	Ber e 1	QMEESPYQTMPRRGMEP
		EGMDESC
		TVVEEENQEECREQMQRQQ
		QQEEMQPRGEQMRR
	Ber e 2	QGSREQEEER
		SQFQGSREQ
		CYNDGD
		NIQRSQKQRG
		ETARKVRGE
		VVVKQA
		AQEPQYRLEAEA
		PAEADFYNPRAG
		FQSMS
		GSREQEEERGRF
		EVWDYTDQQF
		IQNIDNPAE
		DNPAEADFYNPR
		NAPKL
		IREELEQ
		VRGEDDQRG
		QKQRGERYGLRGG
		LRGIPVGVL
		MVVVPQNFVVV
		YEQEELYECRIQ
		LTAQEPQYR
		LWRLNANSVVY
		REELEQQEGGGYNG
		TSPLRGIPV
Black-Walnut	Jug n 1	IDNPRRRGE
		QQSRSGGYDEDNQ
		RGEEMEEMVQ
		EGCQEQIQRQQ
	Jug n 2	CQIQEQSPERQ
		RQYKEQQGRERGP
		KEQQGRERGPEASPE
		ESKGREEEQ
		ERGPEASPRRESKG
		SQSIRSRHESEE
		YFHSQSIRSRHES
		NQDSN
		TPRDRL
		FFDQQEQR
		RFFDQQE
		VIIRAS
		LKSER
		PWGRRSSGGPISLKSE
		SSQSFEDQGRREQEEEEST
		HVSSQ
		FGINGENNQ
		NQLEREA

		TERQSRRGQG
		VPTERQSRRGQ
		RCERQ
		QIQEQSPERQRQC
		KSERPSYSNQ
		AGQRPWGRRS
		QEEEESTG
		EREAKEL
		SRHESEEG
		SQMESY
		PGQVREYYAAGAKSPD
		EEEQQRHNPYYFHS
		PYYFHSQ
		PAGATEYVINQDS
		ERQSRRGQGR
		DQQEQREG
		AGAKSPDQS
		EEIEEIFESQM
		VPHYNSKATV
		SFNMPREE
		NNPGQVR
		GQRPWGRRSSGGPIS
		SYSNQFGQ
English-Walnut	Jug r 1	DIDNPRRRGEG
		RQQSRSGGYDEDNQ
		VRRQQ
		IDNPRRRGEGCRE
		QSRSGGYDEDNQRQHF
		RRRGEGCREQIQRQQ
		RGEEMEEMVQ
		PNECGIS
		YDEDNQRQHFR
		NECGISSQRCEIR
	Jug r 2	SSQSYEGQGRREQE
		LKSESPSYSNQ
		GGPISLKSESP
		CQIQEQSPERQRQC
		GREEEQQRHN
		DDEENPRDPREQYRQ
		SQMESY
		SRHESEEGEV
		NNPGQFREYYAAGAKS
		VPAGA
		ERQSRRGQGR
		DQQEQREG
		PERQRQC
		GAKSPDQ
		YAAGAKSPD
		LDANPNTSM
		EEIEEIFESQM
		REAKEL
		VPHYNSKATV
		PRDPEQRYEQCQQ
		SFNMPREE
		SESPSYSNQFG
		PYYFHSQSIR
	Jug r 3	LRGTVP
	Jug r 4	GGRQQ
		ETFEESQRQSQQGQ
		YNDGS
		NQLDQN
		EQHRRQ
		QQRQQRPGEHGQQQ
		ETARRLQSENDHR
		REEQEREERKERER
		RESESERRQSRRGG
		SERRQSRRGGRDD
		SREEQEREERKERERE
		KERERERESE
		QIPREDA
		NDGSNP
		HSVVYA
		LLDTNNNA
		VIESWDPNNQQF

		AGVIESWDPNNQ
		IEAEAGVIE
		EHGQQQRGLGNN
		PETFEESQRQS
		ERRQSRRGGRDDNG
		KTNENAMVSP
		QRPGEHGQQQRG
		GNPDDEFRPQGQQ
		GCPETF
		HRRSIVR
		ENDHRRSI
		RRLQSENDHRRS
		NDGSNPVVA
		IPREDARRLK
		STVNSHTLPVL
		IPREDARRLKFN
		RGGRDDNG
		QRGLGNNVF
		PQGQQEYEQHR
Hazelnut	Cor a 1	GPVGDKV
		KGGKEKV
		SPFKYVKERVE
		RVEEVD
		GPGTI
		VENVEGNGGPGT
	Cor a 8	RAVNDASRTTSD
		ARASLT
		GVNIPYKISPSTN
	Cor a 9	VGLRRQQ
		EDPQQQSQQ
		GQRQGQGQSQRSEQ
		QQQSQQGQRQGQGQSQ
		CYNDGD
		DEHQRQGQQQFGQRR
		GEQGEQEQQGEG
		DTARRLQSNQDK
		LQSNQDKRR
		RESEQERERQRR
		RPERSRQEWER
		RQGGRGRD
		WERQERQERESEQER
		EWERQE
		LQVVRPERSRQEW
		QVVDDNG
		SREEA
		EQGEQEQQGEGNNV
		QISPL
		GGRGRDV
		KTNDNA
		YNRQE
		NDGDSPVVTVS
		PGCPETFE
		PETFEDPQQQS
		NTVNS
		RGDLQREGLY
		SSSERKRRSE
		IEAEACQIE
		REEARRLKYNRQE
		SSERKRRSESEGR
		NPDDEHQRQGQQQ
		QSNQDKRRNIVKVE
		NTVNSNTLPV
		QQQEGNNVFSG
		IEAEACQIESWDHN
		AQISP
	Cor a 11	PELKKCKHK
		QQEEGNSSEESYGKEQEEN
		QICEEKA
		EHFESR
		LENFTKRS
		YMINRDEN
		VKASREK
		KGSIV
		EQSKGS
		KVFGEQSK
		SQHEEGP
		LLHKHPSQS
		PSREV
		FFFPGPNKQQEE
		GPYYNSRAT
		GAGGEDPESFYRAF
		APGHFEAF
		ERQFDEQQRRDG
		VAPAGHPVAVIAS
		PPRIWPFGGESS
		HEEGPPR
		VREEKRESFN
		PNKQQEEGGRGGRA
		NRDENEKL
		VNEFERDAKE
		EAFYGAGGEDPE
		KHPSQSNQ
		KVRREQLE
		EDPELKK
		NEFERDAKEL
		RDERQF
		RREQLEKV
		GNSSEESYGKEQEEN
		KRESFNV
		PSQSNQFG
	Cor a 14	DIVNQQGRRGESCR
		GSYDGSNQQ
		QQGEMRGEEM
		ITTVDVDE
		EDIVNQQGRR
		YDGSNQQQQQELEQ

The number of epitopic sequences predicted can aid to pin-point what induces the allergenic properties to the various tree-nut allergenic proteins under study.

In addition, with the tendency of patients suffering from allergy due to one tree-nut, often showing sensitivity to other forms of tree-nut as well, a novel methodology was developed in EpiPro1.1 in order to study for cross-reactivity between different allergens. The obtained result is illustrated in Table II.

The result depicts the high probability of cross-reactivity shown by tree-nut allergens. With Hazelnut and Cashew-nut showing similarities in epitopic sequences among five and four other tree-nut species under consideration respectively, it shows if someone who is allergenic to one of any of these species, might also show sensitivity against the other species. Also, as the number of sequences in common increases, the higher is the probability for a person to show sensitivity against the other species. Also, as the number of sequences in common increases, the higher is the probability for a person to show sensitivity to multiple tree-nuts. Hence, from the result we can observe, someone showing sensitivity to hazelnut has a high probability to show sensitivity against Black-nut as well, and vice and versa.

V. Conclusion

20 different allergenic sequences from over 6 different tree-nuts were analyzed using EpiPro1.1, and a total of 362 epitopic amino-acid sequences were identified through a consensus over 6 different web-servers. Cross-reactivity study showed Almonds to have cross-reactivity against Brazil-Nut, Cashew-Nut and Hazelnut; Black-Nut to show cross-reactivity against Brazil-Nut, Cashew-Nut and Hazelnut; Brazil-Nut to show cross- reactivity against Almonds, Black-Nut and Hazelnut; Cashew-Nut

to show cross-reactivity against English- Nut, Hazelnut, Almonds and Black-Nut; Cashew-Nut to show cross-reactivity against English-Nut, Hazelnut, Almonds and Black-Nut and Hazelnut to show cross-reactivity against all 5 other tree-nuts species under study. With 20 different allergenic sequences, the task to predict the consensus among the huge number of possible epitopic sequences obtained through various different web-servers and further analyzing it for cross-reactivity would have been practically impossible and heavily prone to human error. EpiPro1.1 does not only make it simple and time efficient, but also significantly bolsters the accuracy of the obtained result.

TABLE II CROSS-REACTIVITY

Tree-Nut	Cross-Reactive Sequence
Almonds	
Brazil-Nut	DFYNP
Cashew-Nut	RQEGGQ
	QQGRQQ
	QQQQQ
Hazelnut	IPYKISPST
	EQEQQG
	YNRQE
	ETFED
	QQQQQ
Black-Nut	
Brazil-Nut	REQEEE
Cashew-Nut	EESTG
	EAKEL
Hazelnut	RGEEM
	SNQFG
	IDNPRRRGE
	QQSRSGGYDEDNQ
	RGEEMEEMVQ
	EQIQRQQ
	CQIQEQSPERQ
	GREEEQ
	SRHESEE
	YFHSQSIR
	DQQEQR
	GGPISLKSE
	QGRREQE
	ERQSRRGQG
	QIQEQSPERQRQC
	PSYSNQ
	REAKEL
	SRHESEEG
	SQMESY
	REYYAAGAKS
	YAAGAKSPD
	EEEQQRHN
	PYYFHSQ
	ERQSRRGQGR
	DQQEQREG
	GAKSPDQ
	EEIEEIFESQM
	VPHYNSKATV
	SFNMPREE
	NNPGQ
	SYSNQFG
	GGPIS
Brazil-Nut	
Hazelnut	CYNDGD
Cashew-Nut	
English-Nut	RRGEG
	EAKEL
Hazelnut	PESFYRA
	YNSRAT
	PGCPE
	SPVVTV
	QQQQQ
English-Nut	
Hazelnut	RGEEM
	SNQFG
	QSQQGQ
	TARRLQS
	EQERE
	ERESE
	IESWD
	IEAEA
	PETFE NPDDE
	GCPETF
	ARRLK
	GNNVF

REFERENCES

[1] D. Chandrika, International Journal of Otorhinolaryngology and Head and Neck Surgery, 3(1): Jan 2017, pp.1-6.

[2] https://www.foodsmatter.com/allergy_intolerance/miscellaneous/articles/food_allergy_india.html.

[3] A. Johri, Neelabh, M. Srivastava, "A computational study of B-cell epitopes of wheat allergens and identification of its IgE binding residues," Proceedings of the 12th INDIACom; INDIACom-2018; IEEE Conference ID: 42835, pp. 1095-1099, 2018.

[4] J. V. Ponomarenko, M. H. V. V. Regenmortel, "B-cell epitope prediction," In: Gu J, Bourne PE, editors. Structural bioinformatics, second ed. John Wiley & Sons, Inc, pp. 849-879, 2009.

[5] http://ailab.ist.psu.edu/bcpred/.

[6] A. Johri, M. Srivastava, unpublished.

[7] https://www.foodallergy.org/common-allergens/tree-nut.

[8] www.ncbi.nlm.nih.gov.

[9] Y. E. Manzalawy, D. Dobbs and V. Honavar, "Predicting linear B-cell epitopes using string kernels," Journal of molecular recognition, vol 21, pp 243-255, Jul 2008.

[10] Y. E. Manzalawy, D. Dobbs and V. Honavar, "Predicting flexible length linear B-cell epitopes," In Computational systems bioinformatics, vol 7, pp. 121-131, NIH Public Access, 2008.

[11] S. Saha and G. P. S. Raghava, "BcePred:Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties, " In G.Nicosia, V.Cutello, P. J. Bentley and J. Timis (Eds.) ICARIS 2004, LNCS 3239, pp. 197-204, Springer,2004.

[12] J. Chen, H. Liu, J. Yang and K. C. Chou, "Prediction of linear B-cell epitopes using amino acid pair antigenicity scale," Amino acids, vol. 33, pp. 423-428, September 2007.

[13] M. C. Jespersen, B. Peters, M. Nielsen and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," Nucleic Acids Res, 2017 (Web Server issue).

[14] S. Saha and G. P. S. Raghava, "Prediction of continuous B cell epitopes in an antigen using recurrent neural network," Proteins: Structure, Function, and Bioinformatics, vol 65, pp. 40-48, Oct 2006.