

Interview with Professor G.P.S Raghava - discussing Bioinformatics, Research, & science in India

Author : Bioinformatics Review

Categories : [Interview](#)

Date : February 19, 2019



Question: So, more than 200 webservers and 30 databases, and more than 200 research papers in reputed journals. So how do you come up with a team and diverse ideas to implement those?

Answer: *That's an interesting question, I never ever thought that I will publish 200 papers or 200 webservers. After completing my M.Tech. from IIT Delhi, I joined a job because like other persons I also needed the job in 1986 at the Institute of Microbiotechnology, Chandigarh, and I was really happy because I got the Class-1 officer job and I was an odd one out in that institute, everybody was microbiologist because the name of the institute is Institute of Microbiotechnology Chandigarh, so I was hired by the institute to look after the computer center, so I was called as the computer scientist, so initially I felt why they give me the designation of a computer scientist though I am doing the computer service job just to maintain the issued computer to provide service to our users, the biologists. So, during 4-5 years I developed the different types of general software, like Payslip, etc. I realized that the software has a limited half-life because after some time new software will come and the existing software developed by me will be finished. So, this was going on and I thought I need to do something actually. I will do more than what they expect from me. Then in 1990, one student came to me because he was*

working on Immunology, his name was Aveesh and he said he has a problem, he wanted to implement software in immunology. The software was simple, the antibody-antigen concentration from ELISA data, so there was some software already developed for the same. He wanted just to implement it. I saw that paper and after visualizing it, I realized that I can develop a better tool than the tool developed by previous scientists. So, I developed a better tool and tested it experimentally, and we demonstrated that our method is better than the previous method and we sent to the journal, and it got published. So it was basically a wow moment! for a person who has been hired for the service is publishing in high impact factor journal. So I got excitement. After that, I thought whether I will be there or not but my name will be there in the literature and it will be a good contribution to the scientists. So that was the starting point, after that, I just have a simple look at what are the different types of problems that users are facing or bioinformaticians, and I have a lot of friends, I talk to them they give me the problem and in order to help them I write the software and in that I have become a more diverse scientist because I am not working for myself I am working for the community. So, this was the reason why I work in different fields.

Question: Sir, bioinformatics is really an interesting field, how did you get interested in this particular area? Is it like you were actually in biology or computer science together?

Answer: *As I told you, I was hired by the institute to provide computer services, so my background is computers, in fact, I did masters in Physics, after that I did M.Tech in Computer Science, so during my M.Tech I did a lot of programming, so my programming is strong though at the time when I started my career I had nearly nil knowledge of biology, only up to high school like everyone else. That was the point where I applied my computer science skills in biological science only limitation was to know what are the biology problems there and computer in biology is always in demand though in India we know from last 20 years only. I realized by reading literature that it has been important for a long time, even one of the best contribution in India in the field of computational biology was by Prof. G.N. Ramachandran and it was around 1970. He wrote how to compute possible dihedral angles of proteins so that's why its called "Ramachandran plot", where they have defined what are the allowed or disallowed regions and that's one of the best contribution of computational biology in the old times. I thought this field is important so I got into that.*

Question: Sir, you have worked with international organizations like EBI, Oxford, UAMS, so, based on your experience, what suggestions do you have for people working in India become at par with those levels?

Answer: *There are two to three points I would like to mention. One point is the team, in India, as an individual we are as good as anybody in the world, so individually, there is no problem but when the question comes on the team then there is because most of the Indian researchers/ students, they have internal issues in the sense, if you talk to the students, then they will blame their seniors or guides, if you talk to the guides they will blame on students, this is the major problem in India, otherwise, as an individual you are as good as anybody in the world. The second thing is infrastructure, because not now, but in 1996 I was in the Oxford institution for two years. At that time, the internet and other facilities like fiber optics and others were completely there, in India the internet was just started, so they got better facilities than us but now, fortunately, India has better facilities but I am talking about the old time, so there was good infrastructure in all these countries. Another thing is they know that hire and fire policy is there, so what happens if you would work then you get all the advantages and if you don't work then you will get fired. In India, this is not the situation, particularly in the government set up, once you join, then*

whether you are working or not, you are getting the same salary, so the people adopt that. But if we have the feeling that we want to contribute to the country or we want to do something for the community, then these are not the problems, then you can make as good contributions as anybody.

Question: Sir, you have been a significant contributor to the open source community, you have developed many web servers and databases, and you have kept everything free for the users. Although you could have made a fortune over that and you could have commercialized the codes. What inspires you to do that?

Answer: *I will go again back between 1980-1986 when I joined the institute I was providing computer services to the institute. At that time, there was no facility nearly and email only existed and fortunately, in our institute, we established an email facility as anybody like in 1991. This facility was not there in most of the institutes in our country. At that time, how I was getting the sequences in databases from the different organisms to my biologists via email. So, there was EMBL, EBI was not there, in EMBL, there was a repository of free software, where you set an email with the help, it will give the commands you set in the next email, you send the command and you could download the software. Even at that time, you could not download the big software in email but you could send them in the pieces, so the pieces will come, I would compile them and then provide service to the user. At that time, I thought that these diaries like European diaries or USA, they are providing in source code free to the public, why that type of tendency is not there in India?!. In India, we are so possessive, we think that we will live forever actually, and we are not sharing this information with anybody, so that's why we are not growing. We have done a lot of work in the past to be claimed but it is not well-documented and not shared with the public. As a result, what happens, the person who generates the data will die and this data will also die, this is the common problem. So, one thing if you want to grow in life, whatever weakness you see in the system, you should not become a part of that system, you should not contribute to that, so if I become one of them, then (not the part) if I feel that something is wrong, what should I do, I should do something different. At that time, I thought I will not do that thing where I will not allow others to do. So, whatever the logic is there, I will share it, second thing is, why we respect anybody, like in USA, UK, or other because if somebody gives to you, you respect to the person, not the person who is taking from you. The USA has already made Pubmed and other facilities, and Europe also, why India can't do it. We have a lot of restrictions in our own institutes, they ask why you are not charging on anything you have developed a lot of software and providing them freely? I said at least there should be some growth, a person who can contribute to the world so the world can enjoy software servers developed by India, otherwise, when you search on Google we go to the USA, UK, and other sites, how many software servers have you searched that are made in India, so I want to change that trend. And the second thing, freeware has one advantage, as in many developing countries, you cannot afford the costly software, if you can provide the free software equally to the commercial software, the people who do not have enough budget can access the updated software. So, it will save a lot of money of the country, so indirectly, I am saving as I have shown in one calculation, if I have the normal charges, then I have saved the total of more than 800 crores. It's a big amount, it's not visible but hits are there, our software servers are heavily used.*

Question: Sir, what's up with your lab nowadays? What are you working on?

Answer: *I have around 8-10 Ph.D. students who are working in my lab. My research group is different than others, I will tell you why. In most of the cases, guide decides what problem a student is going to work on according to the students work on a particular problem. This is a common sequence, so most of*

the PIs are very focussed and they give their students different problems. This is not the situation in my group because I am not interested in my own career I am interested in solving the biological problems as well as to train Ph.D. students. This is an important thing. So, at first, when anybody is joining my group, then I ask a few questions such as “what you are good at?” and “what are your skills?” and then we see accordingly, what different problems he/she can handle better. This is the reason in my group if I have 10 students working on different problems, one of the students is working on “Probiotics and Prebiotics”. Though this is a new field to me the student is interested, he wants to understand what are the probiotics and prebiotics, so she is working on that. Recently, a girl joined my group as an M.Tech student, she wants to work on the disease, I have never worked on that but in this process, I will also learn with my student. So, we are working on a rare disease caused by lysosomal enzymes. So, a database has developed. Other students are working on different problems such as protein-ligand interactions, immunosuppressive peptides because vaccine development is a major part of our lab, we are working from the last 20 years in the field of immunology. Recently, we have completed a very interesting project on “P-features”, what happened, over the last 30 years, in order to classify/annotate the proteins, predict a portion of a protein, you have to complete the features of the protein like the amino acid composition, all these have been discovered over the years by my group as well as by the other groups. The problem is if a new student will join and learn these tools and it will take a long time, so what we did in the last one month, my whole group worked together and all the possible features in the protein have been calculated by a single software and that software will be available in the form of a webserver as well as the source code. So, anybody who will be working in the field of protein annotation/ protein structure prediction, he can easily use our tools and wouldn't have to reinvent the wheel.

Question: Sir, what do you think are the most interesting areas in bioinformatics?

Answer: *That's an interesting question, in my opinion, all fields are interesting. For me, the main challenge is we have to understand the difference between the biologists, experimentalists, and bioinformaticists because we are not generating our own data, so we cannot discover the new things entirely. In order to work in bioinformatics, you should have sufficient data. Nowadays, we are getting a lot of data particularly in the field of genomics and proteomics, so genomics-based biomarkers, proteomics-based biomarkers have a lot of scopes because a lot of data is available, so we can discover different types of biomarkers. So, this is important, I am not saying that the other areas are not important but this field has more data which is available. If you are going to work in bioinformatics, first you have to see what is your interest, what are the problems for biologists or healthcare problems, once you figure out the problem, then secondly, you have to judge it whether you can do it or not because the problem is not important, what important is whether you have the capability to solve it or not. That's a problem because a lot of people come in the bioinformatics field without judging themselves whether they are capable of developing tools or not. So, that's the important part, you have to judge your strength. After knowing your strength, understand the problem, the next question is whether sufficient data is available or not, even you are highly skilled the problem is there but sufficient data is not there, you can't do anything, because in bioinformatics everything is based on the data, so if data is not there then it would be a problem. Overall, I would say, it more or less depends on the person, whether a particular problem excites you or not, whether you have the capability or not.*

Question: As new technologies are evolving, where do you see a bioinformatician working in 50 years? Does he have a future?

Answer: *Yes, the reason is simple, this biological field, despite of all development and the progress over the years, microarray data, chip-seq data, or RNA-seq data, still we are unable to understand even 1% of our living organisms, we don't know anything, only thing is we are working on pieces and that's why a lot of data has been generated and we don't know what to do with it. So, in future also, a lot of scopes is there and particularly in bioinformatics because more and more data will come but what will you do with that data and nowadays, bioinformatics have become more important than biologists, because biologists can produce the data but even they cannot interpret this data, so they will need bioinformaticians. Bioinformatics has a lot of scopes, only thing is whether you have that capability or not, that's the most important thing. That's the big problem anybody who wants to jump in bioinformatics just because of the scopes, will not be successful, you have to check what is your capability, where you can fit very well, what you can do. It's not that whether genomics is important or proteomics is important, protein structure prediction is important, there is a lot of areas in bioinformatics, you have to see where you fit in. If you are good in computers, you have to take the particular type of problem, if you are good in biology, then you have to take another type of problems, if you are good in chemistry, then you have to take other problems. So, a lot of scopes is there, there is no doubt.*

Question: What method have you found most helpful in training your research staff/team in the use of databases? Which technique have you found quite helpful?

Answer: *I think regarding the learning of the group, frankly, speaking if you go from traditional approach, where one person will teach another will learn, that won't do, so in my group, the culture is learning with each other. I am there but they work in a healthy environment, they are talking to each other, learning from each other, if they are not able to understand, they come to me. So, training human resource development is a major challenge for me, not for database development, for any bioinformatics work because if we do not train the next generations, we will die after some time, then the field will also die. So, I want to make sure the field should grow, so I have different concepts for learning. First thing, it is my record in the old years, if anybody is not getting any pay or face problem in Ph.D. like in my previous institute also, when they are facing problems, so they come into my group, so I am not interested in what I will get from the student, I am interested whether I can make the career of that student or not. So, the trust between me and the student is quite strong. I will provide a healthy environment, they talk to each other, they learn and when we feel that learning is slow, what we are doing we are developing software itself. I gave you the example of P-feature where we have computed all the features of the protein and made it a package which has been stored in the last 20th, so if anyone will start, it will take a long time, but if he will my software, he can learn in few days. So, in the same way, in my group, infrastructure is there, everybody has access even they have access to each other's area of the computer, so, let's say, one webserver will be developed by student X so student Y also has the access, so that they can learn from each other because you can learn faster by examples than the given big lectures. So, openness is helping in training my own students because they are not fighting with each other, not competing with each other, they are happy with each other and at the same time, we are organizing workshops, conferences from last 20 years. If you see the average, I have organized at least two conferences every year. In these conferences, we are giving training of latest trend including databases. For example, earlier they were using MySQL database, nowadays, it is not particularly possible because unstructured data is here, so we are using NoSQL technology such as MongoDB, so we are teaching the students that how to be ready for next-generation databases.*

Question: What is your all-time favorite piece of bioinformatics software and why would you prefer

that?

Answer: *If you are asking about a software developed by others, so I would say BLAST, PSI-BLAST particularly, because PSI-BLAST is not only doing the searching it helps you to generate the evolutionary profiles (psi profiles) and it has been shown in most of the research work if you use an evolutionary profile for prediction than a single sequence, then it will better results. So, PSI-BLAST is one of my favorite tools, I am using it in a number of methods.*

Question: Sir, you have developed a lot of software and tools, which computer language do you use for developing them?

Answer: *I think I use many languages, sometimes, I enjoy to try a different language. When I started my career, the first software was in GW-BASIC, the next was in Pascal, another was in FORTRAN, so I enjoy to write. If you would say a number of programming languages, then I know at least 20 programming languages. So, the important question is in what profession do you use. Earlier, during the initial phase, I used C, I have done a lot of programming in C because a lot of computers supported that, so is best in speed. Then, in 1996, I used Perl in my Oxford group and I thought that most of my work, where I used structure predictions, Perl can do it and probably fast, I am not saying that Perl is a fast language than other programming languages but if you are doing small jobs, Perl is one of the best choices for it. So, for the number of software, I have used Perl. Similarly, nowadays, I am switching more towards Python, the reason is that Python has developed a lot of libraries and even if you are not coding, you can these libraries to implement the machine learning techniques, data mining techniques, so it is easy to use. So, nowadays, we are focussing more on Python.*

Question: So, coming to Python and Perl, big data, AI, the blockchain, as a bioinformatician which one are you focussing on like big data? Do you see any role of blockchain in the coming future for bioinformaticians?

Answer: *Regarding the AI and big data, that's an interesting question, why I am saying interesting because these are different meanings, nothing has changed, I will tell you a good example, nowadays, people are talking about the neural network, particularly, deep learning, we have been using neural network from a long time, maybe 15-20 years back, nothing is great about it, and what they are portraying in the deep learning, we use the cascade network, that part is also doing the same type of job. But now, they are beautiful names/ buzz words as well as more data is there. When earlier enough data was not there, so even if I apply deep learning with multiple layers of a neural network, it will not give good results over optimization. So, in previous methods, I used support vector machines (SVMs). I focus more on the problem rather than the techniques because implementing any technique is not the problem, whatever, you said, AI, etc., it is easy, the only thing you will have to see is what type of technique is required to solve your problem, so for me the most important part is solving the problem. Let's say, as I was talking about Probiotics, a lot of people are coming in the market, nobody knows how is it similar or different it is from the pathogenic bacteria. It may cause a certain disease. So, we are doing half of the technology. So this is important to go through scientific ways and find out what is there in the probiotic strains which are not in the pathogenic strains whether it will change our immune system or not.*

Question: Currently, one of the biggest concerns in bioinformatics is data deluge. A few weeks ago, I read an article published in nature and those people were actually confused about which data to

archive and which data to discard. Because from our point of view, everything is important. So, what do you think what measures should we take or what should we do? And recently, some researchers are trying to reuse the keywords which are already present in the datasets. What measures should we take regarding this?

Answer: *This is a big challenge actually, nobody knows about it because we are processing a lot of data and most of the data is in no use. Even if you see TCGA data, its huge data and unfortunately, the data we need is not there. That's the problem here, we use a lot of cancer genomics data but only limited samples are there. So, the requirement is too high and the existing techniques or storage capacity is not up to the level. We should take care and I think maybe somebody would come and do better mining than the previous ones, it happens in the microarray data, earlier, they were giving you the finished information, then people force that raw data should be there because maybe the new individual is smarter than the previous one, so they may mine better information than the old one. So, imitations are there and I cannot comment whether we should discard it, ideally, we should maintain it.*

Question: *One concern in bioinformatics is that, unlike the software which you developed that are freely available, there is some software which is not available for free and they are charged or over-charged sometimes. So, that also impacts research for the people who cannot afford it. What are your views about it?*

Answer: *This is one big challenge because I have sat in the most number of committees and most of the bioinformatics researchers, they request for commercial software and these are quite costly because they see your pocket rather than the actual cost. So, I am the strong opponent for the commercial software, so I simply, say no you have to use academic software because I am not seeing any commercial software from the algorithmic point of view, which is better than the academic software. Academic software is as good as the commercial software and they are free, you can implement them, the only thing is that lack of awareness is there and there is a problem with users that they are not interested in minimum input and maximum output because we want maximum grants and for them need to show the number figures. But the unfortunate part of this country is that we are not evaluating how much time you will get it and in how much time your paper will get published. The day you will that restriction, they will not go for the commercial software, they will go for the free software. So, I want to give an important message, if you are to become a good researcher in your life, you have to think about why you are doing this. You are doing the research because you want to serve the community but if you are consuming a lot of money from the country unnecessarily, then you are not serving in the country. So, we should think about it and use the minimum budget, and maximum output using the freeware tools and a lot of free data is there.*

Question: *As a Ph.D. student, I have faced a problem that when I am working on a particular project and I need some software, I will search and I will definitely find one but as a Ph.D. student, we should be aware of all these pieces of software because they are very important for bioinformatics, so what do you think how important is it for students to make aware, especially the research scholars, they are not aware of all kinds of software because this is difficult to read each and every issue of the journals, there are many, so what do you think what should we do in this regard and how important is it?*

Answer: *I don't know about others, for me, you give me a topic I will go on Google and I will get it. I try a lot of things, I learn a lot of things like this only. So, if you say any topic, you put proper keywords,*

there are two or three resources, one is to use the wiki, so you will get the introductory sort of thinWikipediaipedia, if you will go to Google, you will get a lot of things, then you go to the Pubmed. If you are in bioinformatics, go to the Pubmed and type the keywords, if you are giving the right keywords, you will get the software, and it will take a few hours only. Only thing is, a major problem I told to my students also, the major part is inhibition if you can overcome all your inhibitions, because when you feel anything new, you have a lot of hesitation, how to go, how to do if you can break that inhibition then it is easy. One thing is important in Indian system particularly, we have a habit of spoonfeeding, for everything we go to the teacher and say “please Sir tell me this”. That habit should be removed at the level of Ph.D. You are fortunate nowadays that you are in this era of internet where you can get all the information, just put the keywords, just for say put ‘ppt’, you will get it, you put the word ‘video’, you will a lot of videos, only thing is you should have the time and that’s all. According to me, that’s not a big thing.

Question: A question that every bioinformatician in India and the world wants to know that how can someone join your lab and what is the criteria that you look for or what are the things that you look for in an ideal Ph.D. candidate or someone joining your lab?

Answer: That’s an interesting question, I was thinking about whether I should answer or not. The reason is that I am not too much worried about the students’ qualifications because whatever the student requirement is there, I will be happy because I feel any student who comes up to this level after doing M.Sc. and all he/she have some talent, only thing is we have to find out what is that talent and how to use it. So, the only thing which we need either the person should have a biology background so that at least we can teach them the computers, there is no problem in that, or a computers person is there, it is also fine, we can teach biology. If both things are not there then it would be difficult. So, this is not my requirement, that’s the requirement of the institute, let’s say if you in M.Sc. then you should qualify first your exam, and then you can apply to our group and we will admit there. Because that’s not my requirement, that is the institute’s requirement. So, I am not that particular that I need that type of student, I have shown the examples in past, where students have been thrown out by some PI because that student is not good, they come to me and within a period of time, they did good research and they are successful. So, I believe in the theory, training the best people and doing the best is easy but taking the people who are not best and making the best is challenging. So, I take it as a challenge, if any student comes to me and wants to do Ph.D., for me that’s a challenge. So, I work as a team rather than as a guide to the students. Let’s say if you come to me, I will take you like my other team member, I will talk to you, I would ask questions, I will take your ideas, I will get some ideas, so, that’s my way of working. Rather than being one source of knowledge and giving to everybody, there is a number of the source of knowledge, and then we are working together, and we are a community, and we can easily solve the problem. Why is this important? Because a number of scientists/ researchers think that they are not good strength, that’s why they cannot do the work, and that’s a good excuse for not to work. So, if you want to work then there is no problem, you have to take it as a challenge because taking the best and further improve it, is the easiest job. But taking a person who nobody is taking and you are making them up, that’s important. For me, that’s a service, frankly speaking, because I am not making my career, I am not interested in my career. My career was completed in 1986 when got the job of Government Class-I officer because I know even if I will not work for my whole life, my job is safe, secured, and everything is there. So, during my whole life, I just worked to provide service to the community. The service is in four or five forms. First is, basically training the manpower, I am providing training, whatever knowledge we are gaining in our group, we are also giving it to others, so the competition will be there. It’s not like that

we set some expertise in our group, it is not available outside, so, whatever expertise we have developed in our group, we have shared it. Second thing, for me, my students are not just to do my research, they are, by associativity or let's say, by service to the nation and I have trained them and they are good in that. So, I am not saying that I am producing the best students, what I am saying is that anybody who comes here, I will improve him, whatever the possibilities are there, and what I am doing, I am just finding out from the student himself that what is his talent, so, I am working with him just to tell him that look this is your talent. You can do the great in this job. That's another service, national recognition is another service. Because I tell students that you are not working for me, you are working for the community and nation, so there is no internal fighting. In the last 20 years, more than 30 students have completed a Ph.D. in my group as well as there are a number of projects, it's not my contribution, its the contribution by everybody in the team, no internal fighting has been there so far among those 30 students. That is the source of our productivity, we are working together, we want to serve the community, that's it!

Question: At last, I want to ask, what is your opinion about bioinformatics? Do you think it is just making the castle in the air or this is just prediction-based or simulation-based, and nothing more we can do with it?

Answer: *For me, it's about solving the problems. Sometimes, there is a misconception that with bioinformatics we can predict anything, no, I consider that bioinformatics will help you to prioritize. Let's say if you are a biologist, you want to work, there is a way, you try randomly, for say, you want to a particular epitope in a protein, one way is you can try all possibilities, if 200 amino acids are there, nearly 192 combinations you will have to try and that will take a lot of time as well as a lot of money. So, me, I will see can I prioritize, can I rank it, rather than all possibilities, try these top 10 first, and I demonstrated it, in the last 4-5 years, we worked on a program, where from bioinformatics, we predicted theoretically that these are separate peptides, experimentally synthesized the chemical, and checked them, and they are the best peptides in the world, where my counterparts the biologists are not able to discover these separate peptides over the years and I discovered them in a few years. So, bioinformatics has a lot of power, I demonstrated that if you combine experimental science and this theoretical science, you can do better. That's why you should see my papers, they are heavily cited, they are not being read by only bioinformaticists, they are used by biologists for prediction. Unfortunately, in India, we don't respect each other's fields, if you are a biologist, they would say, what is in there bioinformatics and vice-versa, that's the problem, that's why they are not collaborating, they are taking the full advantage of our bioinformatics community. I believe in a few years, they will believe more on the bioinformaticians and they will utilize the knowledge for their own experimental work.*

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)
- [Facebook](#)
- [Google](#)

