The question of what test is most accurate is similar to what is accurate in translating the Bible - not only do you have to worry about the language you are translating to, does the meanings match. Maybe it isn't a great analogy but the point I'm trying to make is there really is not a CORRECT answer, there are useful answers and certainly you can have sloppy answers, and even introduce errors at different points. In order to do your question justice we need to parse it into multiple layers.

**The variability of the biome when you sample**

Are you taking an appropriate sample - too big is not good
The buffer is meant to preserve the DNA, it is not perfect but does a good job. If you don't have enough buffer because your sample is too big, this is not good and can introduce "issues" that manifest in the ability to read the sample and sometimes in an unknown way.

Where you sample - when you take the sample? an assumption is the stool is the same in that sample everywhere? Are you mixing it evenly? Does it really matter? Yes it can, but Potentially Not much (likely most cases).

We haven't even addressed the variation of stool over time / day or the fact that you are using it to assess a 4 dimensional problem.. Where it came from (3D) and when.

Okay but so far it hasn't even made it to the lab. The point of the buffer is that it is meant to preserve the DNA so that you can analyze it. It does a good job however it is not perfect. It is a bit controversial but there has been some information that things can deviate / grow . It certainly is light years better than not using a buffer.

https://journals.asm.org/doi/10.1128/mSystems.00199-16

Still these are not addressing your question which involves the differences between labs.
First lets talk about the labs.

Both biomesight and Ombre have their samples processed by a sequencing lab. So when you get angry at them - just know there is only so much they can do at various stages. This is the reality of what it means to provide affordable / fast sequencing for the individual to use / utilize. Lets get it straight the fact they provide raw data has made a lot of the stuff we all do possible - so that needs to be celebrated.

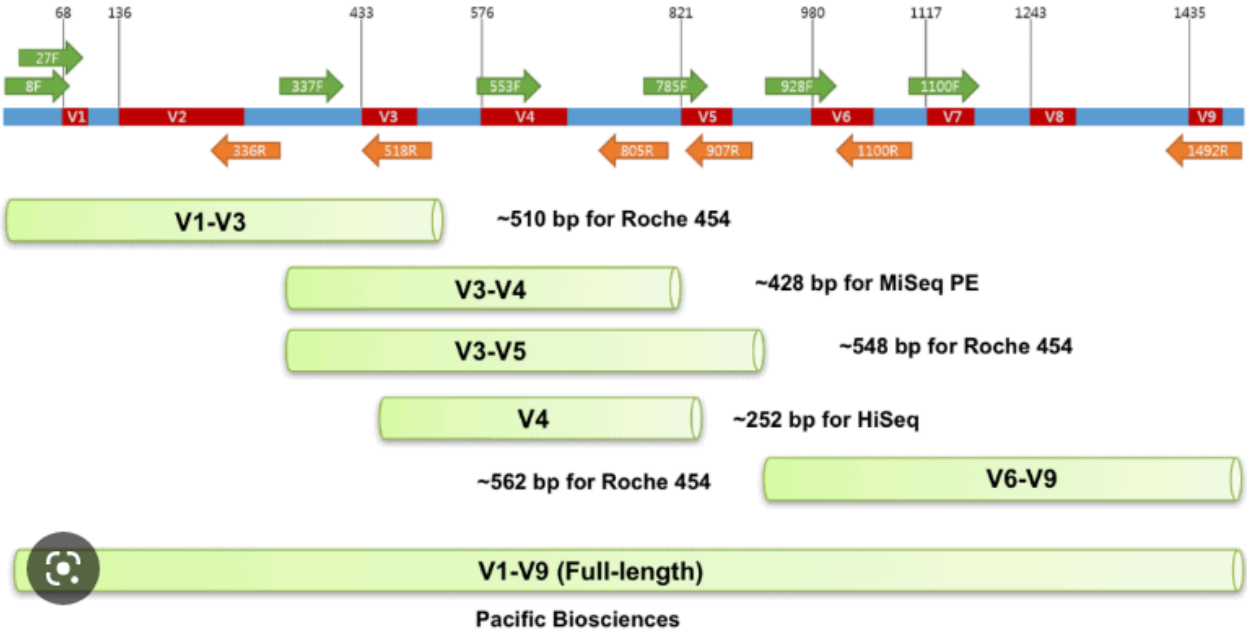
I actually did a video interview with a bio informaticist on the next steps however it is only for practitioner use I do not have the permission to share it publicly.

There are several steps that involve the sequencing process.. I will loosely describe some components - applicable to 16s. There is certain "chemistry" that occurs for 16s rRNA the region of interest that offers conservation and distinction between many microbes is V4, you can

sequence other regions of the 16s gene, you can sequence the whole gene if you like. Some use V3/V4.

The 16S rRNA gene is a segment of DNA that encodes the 16S ribosomal RNA

The 16S rRNA gene is approximately 1500 bp long, with nine variable regions that include conserved regions. These regions of the 16S rRNA gene are frequently used for phylogenetic classification of genus or species level resolving when possible.




sequencing machines are built to sequence double stranded DNA, DNA is usually double stranded in living things. All RNA is single stranded, both messenger RNA and the Ribosomal RNA of the 16S gene.. In order to use the sequence the RNA is reverse transcribed from the single strand of RNA into a double strand of DNA called cDNA or complimentary DNA

https://en.wikipedia.org/wiki/Complementary_DNA

## The four bases that make up sequencing are are adenine (A), thymine (T), cytosine (C)  and guanine (G), but  RNA doesn't contain thymine bases they are replaced with them with uracil bases (U)

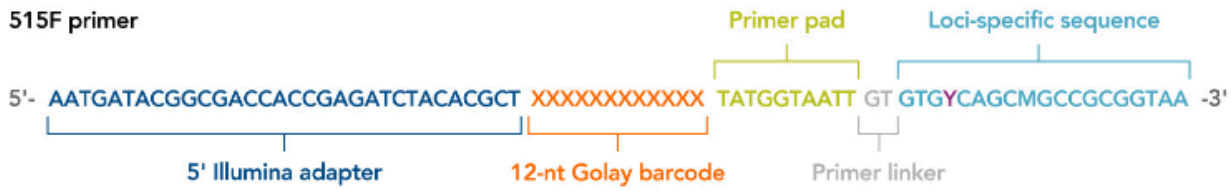
For instance if the original RNA sequence was

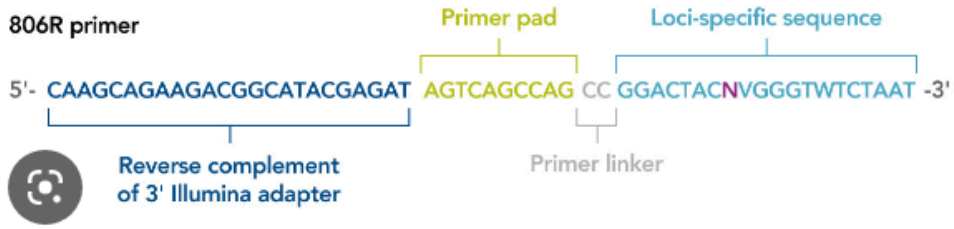
AAUUCG: RNA template where U bases exist instead of T

AATTCG: This matches with the RNA we are amplifying, and it is  now compatible with sequencing platforms

TTAAGC is the reverse is the transcription into cDNA (the compliment)
https://www.sciencedirect.com/topics/neuroscience/reverse-transcriptase

Suffice to say that a certain chemistry processes need to be followed to tag the RNA (primers and adapters) this certainly introduces some variability however there is nothing we can do about it as consumers. Here is an example off some of the chemistry prep used for illumina platforms. There is even variability in how these are implicated for labs using the same sequencing.




https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

The overall process is typically referred to as Amplicon-based next-generation sequencing - and involves tagging / amplification of the reads using PCR

Here is some broad information regarding sequencing - Illumina
https://www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/16s-rrna-sequencing.html A blurb below from Illumina

*“A key benefit of 16S and ITS ribosomal RNA NGS methods is that they provide a cost-effective technique to identify strains that may not be found using traditional methods. Unlike capillary sequencing or PCR-based approaches, next-generation sequencing is a culture-free method that enables analysis of the entire microbial community within a sample.”*

If we are only looking at the 16sV4 region typically 150 BP reading is used forward and reverse - often called pairended 150bp is the read length - that is the length of the reads that occur in 1 swoop if you will.. The amount of nucleotides bases. There are both primers and adapters that tag along the to the tails of the reads on purpose. The primer is the piece that allows you to pick out / hone in on this particular Gene / region.

Typically 515 R and 806 F are primers that are commonly used in this region - since they come along for the ride they need to be filtered out, similarly the barcodes are there to identify your sample - they allow you to parse out RNA reads from others - this later gets demultiplexed, an example of barcode kits
https://store.nanoporetech.com/us/16s-barcoding-kit-1-24.html

You need specific primers to sequence specific regions

Other information that helps to build and understanding

16s has the advantage where the differences between sequences of different microbes are enough to often resolve to species level, but similar enough to categorize UNKNOWN microbes (ones that do not match to a given database), the way that UNKNOWN microbes are match relies on the similarities between microbes at higher levels. For instance something like bacteroides share a lot of mutual information across different species so that when you encounter an unknown bacteroides SPECIES you might not have it in your database but you can with certainty say that it truly is in the bacteroides GENUS.. This is just an example of what could happen. Things can be kicked up to higher levels where there is less mutual information. This type of similarity / similarity does not exist in WGS sequencing - species within the same genus can have significant differences such that identification when your library is incomplete is much more challenging.

The task that sequencing has is to basically READ letters (nucleotides bases) phrases (not full sentences) containing words on a page that tell a story (narrative)- the letters are read at different starting points. Each Phrase is READ left to right. Your read depth is (lets say up to 15 characters and no more).. You are also allowed to READ the phrases backwards… Your task is to put back together the entire story (Narrative) by properly joining where there is overlap the letters that match. This is essentially the Task that you have. In 16s V4 you are looking at about 250 bp of read length needed to reconstruct this region.. So you can read the sequences

FORWARD 150bp and backwards 150BP and get overlap in the middle (you still need to think about trimming primers and barcodes but this is done easily since many tools can automatically do this when the reads are consistent in size). The above analogy is really includes - reading from different books at the same time - you need to filter out the other books (barcodes) and filter out the primers which attach to the phrase ends

So what can go wrong that introduces issues with ACCURACY or maybe Reproducibility, we mentioned some of this already.

If the chemistry isn't right or even if there is, you will not always be able to USE all of your reads. There are inherent uncertainties in the reading process .. the measuring of what nucleotide base you have is done based on measuring of LIGHT.Each base emits a unique fluorescent signal. Examples of how inaccuracies can manifest is that each single read is somewhat dependent on the previous read.. Much like taking a picture behind a DARK BACKGROUND or LIGHT BACKGROUND… the brightness is somewhat affected by surroundings. As the process goes on longer and longer - it becomes more challenging to read reliably with longer reads - almost like trying to count to 1 million and NOT messing up.

https://www.photonics.com/Articles/DNA_sequencing_by_color/a39352

Long story short there are ways to characterize the quality of reads - some of that quality can be impacted by the challenge of sequencing long reads - others by inadequate chemistry where you get chimers https://www.ncbi.nlm.nih.gov/genbank/rrnachimera/ this occurs when things clump together that shouldn't.

Assuming your chemistry is good (process)- andl all the prior things preceding that we talked about we still need to talk about DENOISING - this process basically looks to see what reads are lower quality and removes them from analysis. Keeping lower quality reads will affect your joining .

Denoising
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6087418/

Here is a brief introduction to quality - based on illumina platform.
https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html

You might ask the question, does removing lower quality reads MISS things - well these lower quality issues do not really bias against microbes in the 16s.. . The reason why we can say that is because the lengths of sequences across microbes for this region is very consistent. However if you had a variable length sequencing region like ITS or even WGS - you will have a tougher time joining longer reads because there is LESS overlap to do so.. So quality can somewhat bias against certain microbes in those situations. If you don't remove lower quality reads it will affect the rest of your sample (joining) .. if you don't trim or truncate appropriately.. You can run

into other issues. Luckily 16s has a pretty consistent length so much of these complications are accounted for in multiple stages.

After you DENOISE your sample - you are left with sequences that you can then MATCH against databases.

You are now subject to the available data out there to Determine what your microbe is
There are databases like SILVA / BLAST / GREEN GENES / GREENE GENES II that are collections of sequences and bacteria names / taxonomy. Understand that each of these databases have their own completeness and to some extent their own NAMING convention for the microbes. Even if we know what we are talking about the names of bacteria .. even at higher levels like family have some variations between databases.. There will always be some inherent confusion unless you understand these variations. Keep in mind that the process of bacteria naming is something that involves discovery, isolation, measurement, assignment, and can often result in RECLASSIFICATION or reassignment overtime.

For instance I was handed a list of microbes from a practitioners (clostridium named) almost all of them were reassigned out of the clostridium genus (they are NOT considered clostridium anymore). So someone clearly stepped in shit.. But we are now calling it POO..

Now back to your original question of OMBRE vs BIOMESIGHT
They currently use different sequencing labs (at one time it was the same exact lab)
There is likely some minute differences in the chemistry (however the sequencers are the same)
They have their own pipelines
A pipeline is the process of filtering / denoising / joining / and assigning taxonomy using a database

This last piece is LARGELY  why you SEE differences between them. Using 2 different pipelines Can result in the following differences

Assignment of ASVs - and what confidences are used to assign taxonomy - not everything has complete confidence.

What databases are used to match sequences - their completeness and how they NAME or call microbes
The Denoising process itself - there are parameters

If you know how to compare things - you should get fairly consistent values when comparing between OMBRE / BIOMESIGHT at the genus level.

I probably missed other sources of uncertainty, however I would add this. Despite all that I said these are the best clinical tools (cheap / reproducible / tons of literature / experience and where we have ranges based on). You can use them quite well to solve your issue as long as you know what you are doing and the boundaries of measurement on this platform. I use these tests

to help many people. I have found these tests to be the most helpful in solving dysbiosis and reproducible when things work as intended.

The next answer to your question… it is always best to stay in the same distribution / test.. Compare against others that did the same test. When you step outside of that you start to bark down the path of issues. If you want to do a different test - you should compare within those bounds too. In my experience the variation between ombre and biomesight is minimal but explained largely by the reasons I mentioned above.

***NOTE that I no longer analyze OMBRE testing because they do not collaborate with me in the same way that Biomesight does, they have decided to stop offering my Coupon for ASD children which I had lobbied for (and without announcement to me), and they do not care to work with individuals in this area.***

If you are interested in learning about GUT Balancing - Give a listen
https://youtu.be/BPmdGgXXf8c

Give a read
https://docs.google.com/presentation/d/1ljCSw2dn1f5Qcw6CQybdpHT7O6Sq93Cs2fD2JSSJ-JE/edit?usp=sharing